Ensemble Coding of Crowd Speed Using Biological Motion

Tram T. N. Nguyen (1,2), Quoc C. Vuong (3),

George Mather (4) & Ian M. Thornton (1)

(1) Department of Cognitive Science, Faculty of Media and Knowledge Sciences, University of Malta, Malta.

(2) Department of Criminology, Faculty of Social Wellbeing, University of Malta, Malta.

(3) Biosciences Institute, School of Psychology, Newcastle University, United Kingdom.

(4) School of Psychology, University of Lincoln, United Kingdom.

Word Count: 9916

KEYWORDS: Ensemble Coding; Speed Perception; Biological Motion.

Corresponding authors: Tram T. N. Nguyen, tram.nguyen@um.edu.mt & Ian M. Thornton ian.thornton@um.edu.mt

Abstract

The accurate perception of human crowds is integral to social understanding and interaction. Previous studies have shown that observers are sensitive to several crowd characteristics such as average facial expression, gender, identity, joint attention, and heading direction. In two experiments, we examined the ensemble perception of crowd speed using standard point-light walkers (PLW). Participants were asked to estimate the average speed of a crowd consisting of 12 figures moving at different speeds. In Experiment 1, trials of intact PLWs alternated with trials of scrambled PLWs with a viewing duration of 3 s. We found that ensemble processing of crowd speed could rely on local motion alone, although a globally intact configuration enhanced performance. In Experiment 2, observers estimated the average speed of intact-PLW crowds which were displayed at reduced viewing durations across five blocks of trials (between 2500 ms and 500 ms). Estimation of fast crowds was precise and accurate regardless of viewing duration and we estimated that 3-4 walkers could still be integrated at 500 ms. For slow crowds, we found a systematic deterioration in performance as viewing time reduced and performance at 500 ms could not be distinguished from a single-walker response strategy. Overall, our results suggest that rapid and accurate ensemble perception of crowd speed is possible, although sensitive to the precise speed range examined.

Introduction

As social beings, our ability to accurately perceive and interpret the behaviour of others is crucial to the success of our interactions. Decades of research has shown that our visual system is extremely sensitive to the information contained in dynamic human faces and bodies (Hu et al., 2020; Johnson & Shiffrar, 2013; Knoblich, 2006; O'Toole et al., 2002; Yovel & O'Toole, 2016). Recently, a growing number of researchers within the field of biological motion (Johansson, 1973; for a review see Blake & Shiffrar, 2007; Pavlova, 2012), have shifted focus from studying single, isolated actors, to consider the interactions of dyads (Boker et al., 2011; de la Rosa et al., 2014; Georgescu et al., 2014; Kaiser & Keller, 2011; Neri et al., 2006), groups (Bolling et al., 2013), and crowds (Elias et al., 2017; Florey et al., 2016; Sweeny et al., 2012, 2013; Sweeny & Whitney, 2014; Whitney et al., 2014).

This shift in focus raises new questions about the visual processes that support the perception of crowd behaviour as opposed to individual actors. The goal of the current paper was to determine whether human observers are able extract the average walking speed of a crowd using displays of standard point-light walkers (PLWs; Johansson, 1973). It is known that the speed at which an individual actor moves can convey information about their intentions, temperament, and emotions (Grayson & Stein, 1981; Michalak et al., 2009; Pollick et al., 2001; Troje, 2002). The speed at which crowds of people move may thus offers important information about their characteristics and collective behaviours (Cohen et al., 2008; Moussaïd et al., 2010).

To examine the perception of crowd speed, we made use of concepts and techniques from the ensemble perception literature. Ensemble perception refers to the ability of our visual system to perceive the central tendency and dispersion of clusters in visual scenes with great efficiency, including the mean and variance (for a review see Alvarez, 2011; Whitney & Yamanashi Leib, 2018).

Ensemble coding has been shown to be robust for basic visual features such as motion direction and speed (Atchley & Andersen, 1995; Watamaniuk & Duchon, 1992; Watamaniuk & Heinen, 1999; Watamaniuk & McKee, 1998; Williams & Sekuler, 1984), orientation (Dakin & Watt, 1997; Parkes et al., 2001), colour (Maule & Franklin, 2016; Ward et al., 2016; Webster et al., 2014), brightness (Bauer, 2009), spatial centroid of targets and distractors (Alvarez & Oliva, 2008), and size (Ariely, 2001; Chong & Treisman, 2003). It has been suggested that ensemble-like mechanisms are the basis of our rich visual perception including texture

perception and/or rapid scene gist perception (Alvarez, 2011; Brady et al., 2017; Landy, 2014; Whitney et al., 2014; Whitney & Yamanashi Leib, 2018).

Ensemble coding is also thought to operate on higher level features, such as animate and inanimate objects (Khayat & Hochstein, 2019; Yamanashi Leib et al., 2016), human facial features (de Fockert & Wolfenstein, 2009; Florey et al., 2016; Haberman & Whitney, 2007; Peng et al., 2019; Sweeny & Whitney, 2014), and biological motion (Sweeny et al., 2012, 2013). In terms of crowds, it has been shown that human observers can quickly and accurately estimate the average crowd emotion and gender (Haberman & Whitney, 2007), head rotation and/or eye gaze (Florey et al., 2016; Sweeny & Whitney, 2014), and walking direction (Sweeny et al., 2012, 2013). Synchronisation of crowd movement has also been shown to improve the precision of ensemble judgements relative to when actors behave independently (Elias et al., 2017). The ability to rapidly and accurately perceive crowd characteristics would make evolutionary sense, considering that social gatherings and public participation form an important part of the human society (Massey, 2002) and under certain circumstances we need to make rapid judgment of crowd behaviours.

From a theoretical perspective, biological motion is an interesting case to study in the context of ensemble perception. While, as just noted, ensemble processing is robust at different stages of the visual hierarchy, research has shown that there is little correlation between low-level ensemble representations and high-level ones (Haberman et al., 2015; but see Florey et al., 2016). However, this apparent independence of ensemble perception across levels of visual analysis (Whitney et al., 2014) is almost certainly due in part to the fact that the discrete visual categories examined have not been closely related (e.g., colour and orientation for low-level features, facial expression and identity for high-level features).

More generally, it seems likely that there are ongoing interactions between stages of visual processing (Delorme et al., 2004; Hochstein & Ahissar, 2002). Examining ensemble perception of biological motion can thus offer an interesting opportunity to probe for such interactions, as biological motion is known to involve both low-level (Giese & Poggio, 2003; Mather et al., 1992; Thornton & Vuong, 2004; Troje, 2002) and high-level (Cavanagh et al., 2001; Thompson & Parasuraman, 2012; Thornton et al., 2002) processing mechanisms.

Previously, Sweeny et al. (2013) used biological motion to study ensemble perception of crowd orientation. Their observers were able to rapidly estimate the average heading of intact-PLW crowds. However, heading estimation of scrambled-PLW crowds was at chance-

level. In the scrambled condition, the initial position of each PLW's signal dots was spatially shifted which disrupted the global motion but preserved the local motion. This finding suggests that observers relied on higher-level representations of the PLWs' global motion for estimating the average heading. Poor performance of scrambled-PLW crowds, however, could have been due to the fact that orientation information is unlikely to exist for local signals from individual PLW figures, except for the wrists and ankle joints (Cai et al., 2011; Mather et al., 1992).

This is unlikely to be the case for judgements involving speed. That is, velocity estimates involving simple, isolated stimuli, such as moving lines (McKee & Welch, 1985), as well as more complex random dot fields (Festa & Welch, 1997; Watamaniuk & Duchon, 1992), are known to be both rapid and accurate. Thus, the local motion signals contained within scrambled PLWs could well be sufficient to support the perception of crowd speed.

Of course, this does not rule out a contribution from the global configuration of walkers within the crowd. Ueda et al. (2018) reported that the global configuration of single PLWs did enhance speed detection and discrimination decisions. Furthermore, observers can reliably judge the walking speed of individual PLW figures, even though low-level visual cues such as retinal size and speed vary unpredictably (Mather & Parsons, 2018). It has also been suggested that temporal summation times for intact biological motion are much longer than for simple motion stimuli, as the former relies on estimates involving coherent step cycles, rather than absolute stimulus duration (Neri et al., 1998). It remains to be seen whether such global factors – necessarily involving the temporal integration of higher-level representations -- are also able to influence estimates of average crowd speed.

In Experiment 1, we specifically examined whether global processing of biological motion contributes to the perception of crowd speed estimates by contrasting performance on displays containing either intact or scrambled PLWs. Observers were given 3 seconds to view a crowd consisting of 12 PLWs walking at different speeds. They were then presented with a response array which consisted of 12 PLWs whose walking speed was arranged in ascending order, from slow to fast. Their task was to choose the figure that best matched their estimate of the average speed in the first interval. Blocks of intact PLWs alternated with blocks of scrambled PLWs. We analysed both the accuracy and precision of speed estimates, determining whether there was evidence of averaging and contrasting performance on intact and scrambled trials.

In Experiment 2, we examined whether the perception of crowd speed could be accomplished with reduced viewing time. The three second duration used in Experiment 1could make it possible for observers to sequentially sample walkers. Such serial sampling strategies (Myczek & Simons, 2008) have been proposed as alternatives to the idea that ensemble processes occur automatically and in parallel across the visual field (Chong & Treisman, 2003; Treisman, 2006). Observers viewed intact-PLW crowds under gradually reduced viewing duration across blocks of trials. In the last block, observers had only 500 ms to view the crowds. We sought to establish whether there was a systematic effect of stimulus duration on crowd speed estimates and whether averaging would still be observed for very brief durations. We also carried out simulations to estimate the number of integrated PLWs at both long and short durations.

Experiment 1

Crowd speed estimation of intact and scrambled biological motion

Our main goals in this experiment were (i) to determine whether observers could estimate the average walking speed of a crowd; (ii) to assess whether such estimates rely on the integration of more than one walker; and (iii) to examine the contributions of local and global processes in the context of crowd speed estimation. Observers viewed a crowd consisting of 12 intact or scrambled PLWs for three seconds. They then selected the best matching speed in a response array of 12 speed options. As previous studies have shown that compulsory averaging of motion signals occurs early on in the visual hierarchy (Watamaniuk & Duchon, 1992; Watamaniuk & Heinen, 1999), we predicted that observers would be sensitive to the average crowd speed even when the PLWs were scrambled. With regards to the role of global motion configuration, previous biological motion research has often demonstrated global advantages (Bertenthal & Pinto, 1994; Pavlova & Sokolov, 2000), and the perception of speed in particular, using single PLWs, has been shown to improve when figures are intact (Ueda et al., 2018). Thus, we predicted that estimation of crowd speed would also be more accurate and reliable for intact-PLW crowds.

Method

Participants

Sixteen observers (M_{age} = 25.4 years, SD_{age} = 5.8 years; 8 females; 12 right-handed) were recruited from the University of Malta research community. Sample size was established prior to data collection. The effect size (Cohen's *d*) of statistical tests for the difference in speed

estimation between intact and scrambled PLWs was between 1.6 and 3.7 based on the findings of Ueda et al. (2018) and a pilot study conducted in our lab. We conducted a *priori* power analysis using these estimates in G*Power 3.1.9.4 (Faul et al., 2007) with an assumed power of 0.8 and an alpha of 0.05. This analysis indicated that a sample size between 3 and 5 participants would suffice to ensure statistical power to detect stable differences. We chose a conservative sample size of sixteen due to the introduction of a crowd context and to match previous ensemble studies with non-experienced observers (e.g., de Fockert & Marchant, 2008; Maule & Franklin, 2016). All observers had normal or corrected-to-normal vision, no colour deficiency, and were not experienced psychophysical observers. Observers gave their written informed consent before participating in a single experimental session. A session lasted approximately 45-60 minutes. Observers were monetarily compensated. All methods and procedures conformed to the Ethics and Data Protection Guidelines of the University of Malta.

Apparatus

The stimulus, the task, and data collection were programmed and executed in MATLAB using the PsychToolbox extension (version 3; Brainard, 1997; Kleiner et al., 2007; Pelli, 1997) on a Macbook Pro running OS X 10.10. Experimental scripts and demonstration movies are available on the OSF page associated with this paper at https://osf.io/5j4qe/.

The laptop was connected to a BenQ XL2420-B monitor which had a visible area of 57 x 51 cm, a resolution of 1900 x 900 pixels, and a refresh rate of 60Hz. Observers were seated in front of the monitor at approximately 80 cm viewing distance, using a computer mouse to register responses, and a standard USB keyboard to switch between trials. The experiment took place in a sound-proof experimental booth with no background or overhead lighting to reduce glare.

Visual Stimuli

Each trial consisted of two intervals (Figure 1, see online materials for dynamic versions). In the first interval, a probe array was presented which showed a crowd of 12 green figures arranged in a 3 x 4 invisible grid occupying a 19.5cm x 14cm screen area (subtending 13.9˚ x 10˚ visual angle) for three seconds. In the second interval, a response array was presented which contained 12 white figures which were arranged in the same spatial grid as the probe array, and had increasing speed values from the slowest (top left of the grid) to the fastest speed (bottom right of the grid). The response array stayed on screen until participants

responded. The figures in both probe and response array were intact PLWs on some trials or scrambled PLWs on other trials (see more details below for the creation of these figures).

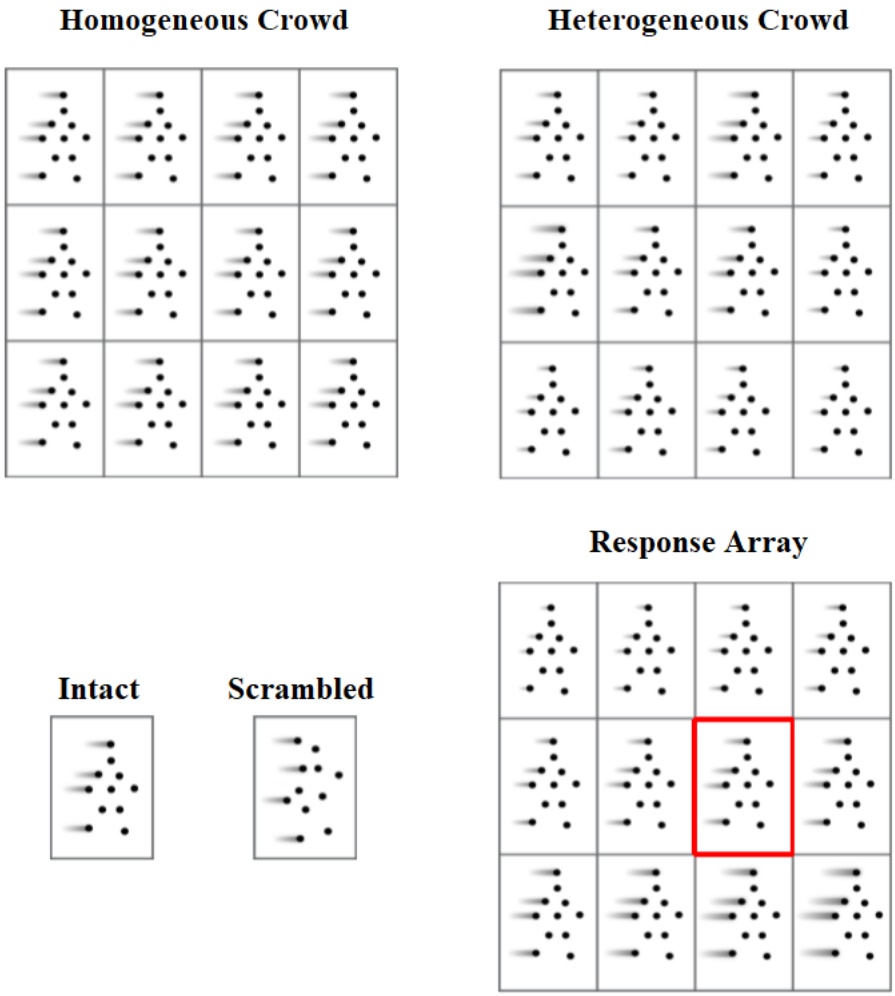

Figure 1. An example of a given trial in Experiment 1. A probe interval shows a homogeneous crowd (top left) in the first condition and a heterogeneous crowd (top right) in the second condition. After viewing the probe, observers select their response in the second interval showing the response array (bottom right). Trials of intact PLWs are alternated with trials of scrambled PLWs within a condition (bottom left). Magnitudes of speed lines indicate different velocities.

We chose to use a fixed response array, rather than a single adjustable figure, for several reasons. First, it maintained the overall display structure between the two intervals, which we felt was less disruptive than switching back and forth between multiple and single figures. Second, it provided a simple way of having the entire speed range constantly visible, and the slowest and fastest speed anchor points available for reference during the decision phase, which we felt might be advantageous. Third, we felt that a single adjustable test figure would be more likely to give rise to speed adaptation effects if participants paused or delayed their responses in one part of the range, given that there were no time constraints placed on responding (Mather

et al., 2017; Mather & Parsons, 2018). Of course, there may also be advantages to using a single, adjustable response figure, such as the ability to expand the range of possible selections beyond the probe endpoints and/or having continuous rather than discrete response intervals. Also, by presenting the full speed range in our fixed probe array, we are introducing an additional crowd, with its own mean and variance, which could potentially interfere with the assessment of the target crowd. Although the parameters of this response crowd are constant across trials -- and thus we felt it unlikely to bias responding -- such issues would be avoided by using a single response figure. Overall, given the above, directly comparing the two types of response method may prove to be an interesting avenue for future studies.

Range of walking speed

According to a recent review of human locomotion speed (Tudor-Locke et al., 2018), walking speed was considered very slow at < 60 steps/min, slow at 60-79 steps/min, medium at 80-99 steps/min, fast at 100-119 steps/min, and very fast at > 120 steps/min. We made use of 12 walking speeds ranging from 40 to 150 steps/min with a fixed increment of 10 steps/min. We coded these speed values as 1 to 12 internally in the task algorithm, with 10 steps/min representing one speed unit.

Intact PLWs

The individual PLWs were created using a motion capture file taken from the database of Vanrie and Veirfaillie (2004). Each PLW consisted of 13 dots (each dot subtended approximately 0.1° x 0.1°) representing the head, shoulders, elbows, wrists, hip, knees and ankles. All dots were always visible, even when they would have ordinarily been occluded by other parts of the body. Each PLW was orthographically projected so that it subtended approximately 4.7° visual angle in height and 2.5° in width at the widest part of the step cycle regardless of their position within the display grid. When set in motion, the original PLW displayed a full walking cycle (i.e., two steps) in one second with an animation rate of 30Hz (30 frames/sec), or 120 steps/min. The slowest walking figure (40 steps/min) required 90 frames/sec to completely animate a walking cycle, while the fastest walking figure (150 steps/min) required 24 frames/sec. We created PLWs with varying speeds by applying an interpolation function to the original PLW's x-y-z coordinates to create new animation frames according to their speeds.

All PLWs walked in place (i.e., like on a treadmill) in the same heading direction, with starting position of the first step being independently randomised for each walker in the crowd.

At the beginning of each trial, we randomised the heading direction between 0˚ and 360˚ about the vertical axis (0˚ = left facing; 180˚ = right facing). The decision to randomise the heading direction, rather than to systematically vary this parameter across trials, meant that in the current experiment we were not be able to fully explore the relationship between ensemble estimates of heading direction and speed. As already noted in the Introduction, observers can rapidly estimate the average heading of a crowd and such estimates are particularly precise for directions approaching the observer (Sweeny et al., 2012, 2013). For speed judgements, however, we might anticipate that sideways views, rather than approaching views, are more informative, as they contain more clearly visible ankle/wrist movements (Cai et al., 2011; Chang & Troje, 2009; Mather et al., 1992). This raises the possibility of an interesting interaction between these two factors. However, as our experimental design already involved a factor with multiple levels (i.e., speed), we decided not to prioritise this question in the current study by systematically varying heading direction. Nevertheless, in a supplementary heading direction analysis, we used a post-hoc binning technique to separate trials at each speed into either roughly forward/backward (0-45°) or side (46-90°) facing views in each quadrant, and we briefly refer to these preliminary findings in the General Discussion. Clearly in future studies, possibly using a smaller range of speeds, it may be interesting to explore the interaction between these two factors more precisely.

Scrambled PLWs

Scrambled PLWs were created using a two-phase procedure. First, the vertical positions of each dot in the intact walker were shuffled, so that, for example, the hip dot might be shifted to the vertical position of the head and the left knee dot might appear at the height of the wrist dot. This ensured that the basic visual extent occupied by the scrambled and intact walkers were the same. Randomly assigning the X/Y coordinates of dots would change the dot density and dimensions. The limitation of using this "shuffle" approach, is that some scrambled stimuli continued to convey a global layout that could be construed as humanoid, if not human. The second manipulation was to randomly assign the starting phase of each dot separately. This breaks the coupling between arm and leg swings, for example. For the scrambled PLWs, the basic speed manipulation, walking phase, and orientation procedures remained the same as described above for intact PLWs.

Task and Procedure

The experiment consisted of two conditions that were completed in a fixed order, the first involving homogeneous (practice) crowds and the second involving heterogeneous (experimental) crowds. Each condition contained blocks of intact PLWs alternating with blocks of scrambled PLWs. The figure type (i.e., intact or scrambled PLWs) in the first block of each condition was counterbalanced across observers.

The homogeneous condition was used to familiarize observers with the probe array and response methods used in the main experiment. Data from this condition was not used in the main analysis, but is included in the online Supplementary Materials. During these trials, the probe array displayed 12 figures walking at the same speed. The speed on a given trial was randomly chosen from a uniform distribution of the established 12 speed values described above. There were a total of 120 trials (12 speeds x 2 figure types x 5 repetitions), with each block of intact- or scrambled-PLW crowds containing five trials.

After a short break, observers proceeded to the heterogeneous condition where the probe array displayed 12 figures walking at different speeds. Observers were explicitly instructed to try and quickly 'guess' the average speed, rather than to sample each figure in turn in an attempt to calculate the average in an effortful manner.

At the start of each trial, we first selected the mean speed of the crowd uniformly among 1-12 speed values. We then generated individual speeds from a normal distribution centred at the selected mean with a standard deviation of 3 speed units. We capped a margin of 0.25 speed units around the requested standard deviation so that array variability remained comparable across trials and observers. Individual requested speeds were also controlled to be between 1-12 speed values. Note that with the standard deviation between 2.75 and 3.25, and individual speeds between 1 and 12, the mean of the requested array tended to fall between 4 and 9 speed units (see crowd speed distribution in Supplementary Materials). That is, the presented mean speed was very rarely 3 or 10 speed units, and almost never occurred for extreme values smaller than 3 or larger than 10. For this reason, we excluded trials with extreme mean speeds in the data analysis, making the actual speed range of 4 to 9 speed values (or 60 to 120 steps/min) in the main analysis. There was a total of 240 trials (12 speeds x 2 figure types x 10 repetitions), with each block of intact- or scrambled-PLW crowds containing 10 trials.

Data Analysis

We analysed estimation performance in the averaging task (i.e., the heterogeneous condition) in terms of both *accuracy* and *precision.* To assess estimation accuracy in this

context, we analysed perceived stimulus speed relative to actual crowd speed (Figure 2). Perceived crowd speed is calculated by taking the average of the absolute response values across repeated trials of the same stimulus speed. In terms of estimation precision, we calculated response variability by taking the standard deviation of the distribution of selection errors (i.e., the differences between true and estimated values). Response variability thus reflects the reliability/precision of an estimate (Figure 2).

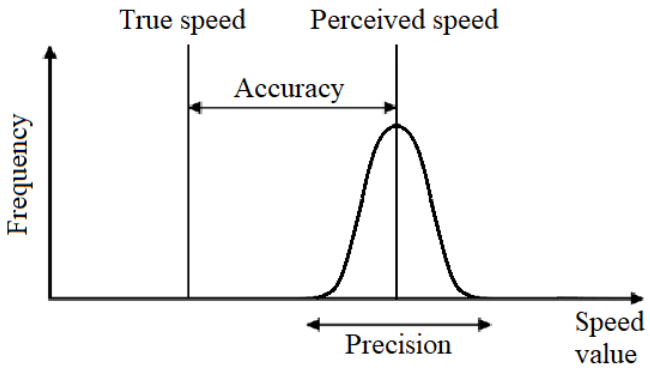

Figure 2. Measurement of accuracy and precision in the current speed context.

We note that while some researchers have used response variability/precision to assess ensemble perception (Sweeny et al., 2013), others have opted to analyse perceived stimulus values or accuracy (Ji & Pourtois, 2018; Marchant et al., 2013; Maule & Franklin, 2016). Considering one property over the other can lead to conflicting conclusions because observers can be precise but not necessarily accurate and *vice versa.* Thus, we chose to analyse both perceived crowd speed and response variability to obtain a more comprehensive overview of data patterns.

We first examined the accuracy and precision of crowd speed estimation using 2 (Figure Type: Intact/Scrambled) x 6 (Speed) repeated measure ANOVAs. Separate ANOVAs were conducted for perceived crowd speed and response variability. We then examined whether there was any specific evidence of averaging by comparing observed response variability to simulated performance based on a random single-walker response strategy. This simulation ran 1000 samples, each of which consisted of 120 trials generated using the exact experimental methods to generate heterogeneous crowds (making a total of 120,000 simulated trials).

Data Availability

The data and analysis routines relating to this experiment can be accessed via the associated OSF page at https://osf.io/5j4qe/.

Results

Figure 3a shows perceived crowd speed as a function of actual crowd speed. It is immediately clear that observers were able to adjust their estimates to follow the true crowd speed. Visual inspection also suggests a possible systematic bias whereby observers tended to overestimate slower speed and underestimate faster speed. We return to this latter pattern in more detail in the General Discussion.

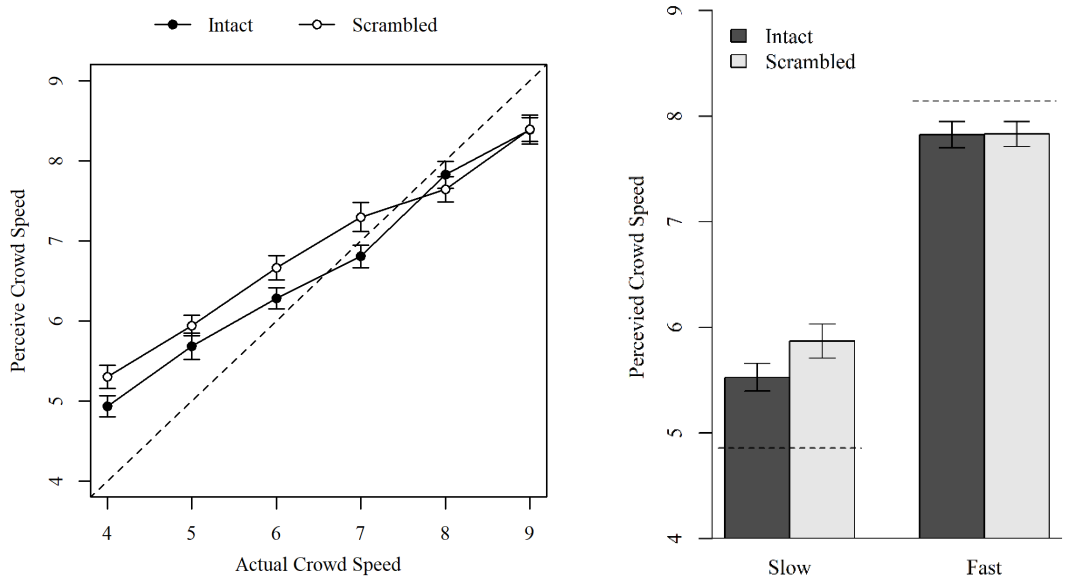

Figure 3. A) Perceived crowd speed as a function of with actual crowd speed. The dashed diagonal line indicated veridical estimation. B) Perceived crowd speed of slow/fast grouped data. Horizontal dashed lines indicate true crowd speeds. Error bars are ±1 standard error.

The 2 (Figure Type) x 6 (Speed) ANOVA showed a significant main effect of Speed, $F(2.25, 33.76) = 135.26$, MSE = 0.78, $p < 0.001$, $\eta^2 = 0.90$. Post-hoc tests revealed that all pairwise comparisons were significant (all ts > 5.12 and all ps < 0.01), confirming that estimated crowd speed varied accordingly to true crowd speed.

The main effect of Figure Type was not significant, $F(1, 15) = 2.82$, MSE = 0.82, $p = 0.11$, $\eta^2 = 0.16$, but there was a significant Figure Type x Speed interaction, $F(3.10, 46.53) = 3.14$, MSE = 0.28, $p = 0.03$, $\eta^2 = 0.17$. None of the pairwise comparisons remained significant after correction. However, the nature of the interaction appears to be clear in Figure 3b, which divides the speed range into slow (4-6) and fast (7-9) trials. To further explore this pattern, we

conducted a follow-up 2 (Speed) x 2 (Figure Type) ANOVA. There was again a non-significant main effect of Figure Type, $F(1, 15) = 1.99$, MSE = 0.24, $p = 0.19$, $\eta^2 = 0.12$; and a significant effect of Speed, $F(1, 15) = 184.02$, MSE = 0.394, $p < 0.001$, $\eta^2 = 0.93$. Importantly, the simple Figure Type x Speed interaction was also significant, $F(1, 15) = 10.23$, MSE = 0.44, $p = 0.006$, $\eta^2 = 0.41$. This clearly shows that at fast speeds estimation accuracy for intact- and scrambled-PLW crowds was comparable, while at slow speeds, observers estimated intact-PLW crowds more accurately (closer to veridical) than scrambled-PLW crowds.

Figure 4 plots response variability as a function of figure type and actual crowd speed. There was a significant main effect of Figure Type, $F(1, 15) = 5.54$, MSE = 0.09, $p = 0.033$, $\eta^2 = 0.27$, no main effect of Speed, $F(5, 75) = 0.85$, MSE = 0.10, $p = 0.516$, $\eta^2 = 0.05$, and a significant Figure Type x Speed interaction, $F(2.92, 43.75) = 2.88$, MSE = 0.09, $p = 0.048$, $\eta^2 = 0.16$, suggesting global configuration of PLW influenced response variability but the effect depended on crowd speed. We ran pairwise comparisons between intact- and scrambled-PLW crowds at each speed point. The results showed that response variability was comparable between intact- and scrambled-PLW crowds when crowd speeds fell within 4-7 speed ranges (from 70 steps/min to 100 steps/min; ts < 1.09; ps > 1.74). When the average crowd speed was faster than 100 steps/min, the presence of the intact global configuration benefited the reliability of crowd speed estimation (ts > 3.10; ps < 0.05).

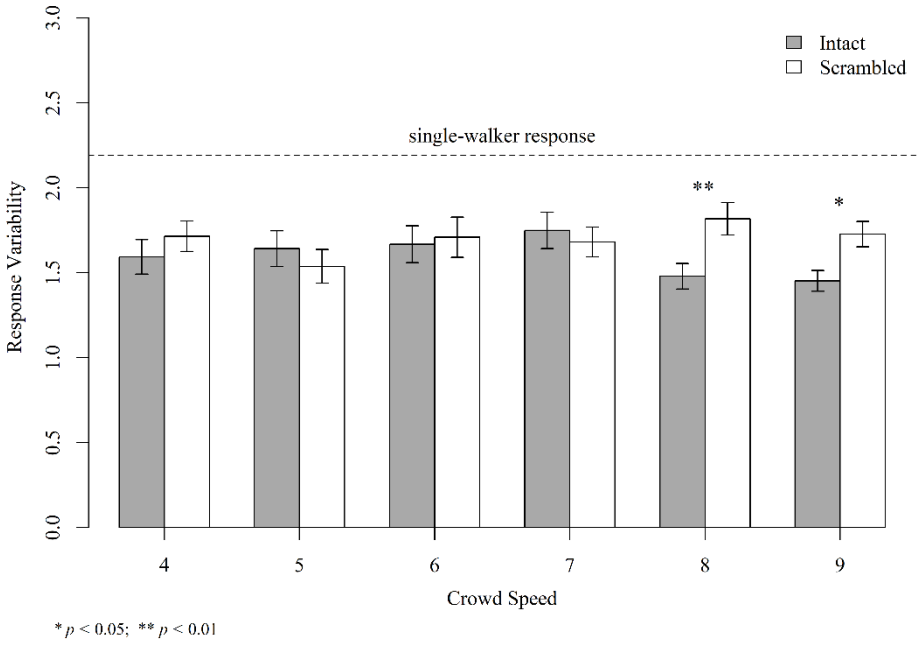

Figure 4. Response variability as a function of figure type and actual crowd speed. Error bars are ±1 standard error.

The simulation of a single-walker response strategy showed that response variability fluctuated within a distribution with a mean of 2.69 and a standard deviation of 0.16. We chose the cut-off value at 2.19 which was more than three standard deviations from the mean (indicated by the horizontal dashed line in Figure 4). We compared response variability of the heterogeneous condition with this value beyond which would indicate that observers based their estimate on one walker of the crowd and the average crowd speed was not coded as ensembles.

The result of one-sample t-tests showed that estimation of the average crowd speed for both intact- and scrambled-PLW crowds was significantly better than single-walker simulations, regardless of the actual crowd speed (intact: ts > 4.10, ps < 0.001; scrambled: ts > 3.91, ps < 0.01). Thus, observers appear to have based their estimation of crowd speed on more than one walker.

Discussion

In Experiment 1, observers estimated the average speed of a crowd consisting of 12 intact or scrambled PLWs using a 3 second viewing duration. We found that observers perceived crowd speed in accordance with the actual mean speed and were able to integrate more than one crowd member to compute their estimates. Further, we found that ensemble perception of crowd speed could rely on local motion alone, consistent with previous studies demonstrating that motion averaging mechanism operates at early stages of visual processing (Watamaniuk & Duchon, 1992; Watamaniuk & Heinen, 1999). However, crowd speed estimation for faster speeds became more precise with the presence of globally intact human configurations. In terms of the accuracy of speed estimates, there was also an advantage for intact figures although this was limited to slower speeds. These findings are consistent with the general biological motion advantage for intact figures (Bertenthal & Pinto, 1994; Pavlova & Sokolov, 2000) and speed perception research using single PLW (Ueda et al., 2018).

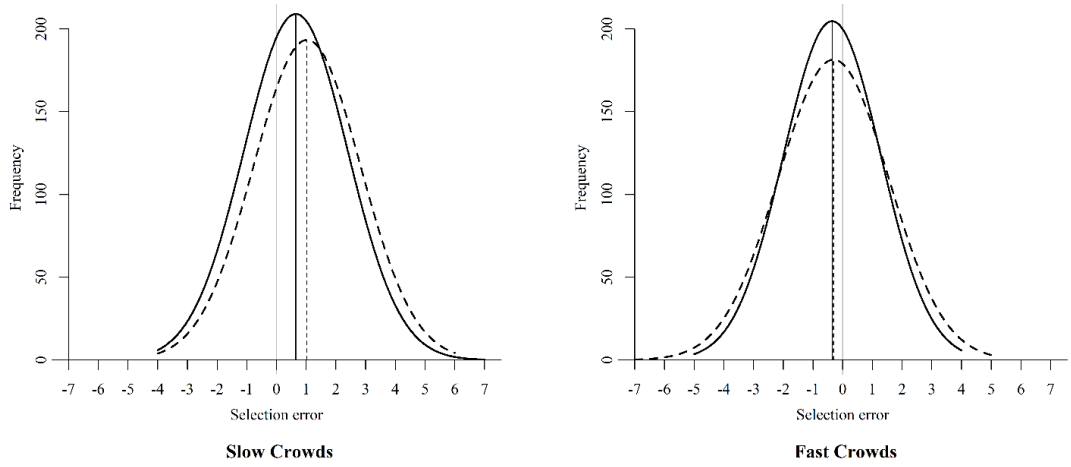

Figure 5. Distributions of selection error for intact-PLW crowds (solid curves) and scrambled-PLW crowds (dashed curves) in the heterogeneous condition. Global configuration of PLWs enhances estimation accuracy in slow crowds (left) and estimation precision in fast crowds (right).

For both of our dependent measures, then, the beneficial effect of a global configuration appeared to depend on actual stimulus speed. As illustrated in Figure 5, the global configuration of PLWs was beneficial for slow crowds in terms of estimation accuracy (left panel), while the precision of speed estimates of intact-PLW crowds was better for intact figures when the crowds were fast (right panel). By examining both measures, we have shown that measure of precision and accuracy can provide useful, complementary information to understand different aspects of observed performance, such as highlighting the differential effects of global advantage for slow and fast crowds in our task.

Experiment 2: Crowd Speed Estimation with Reduced Viewing Duration

One of the hallmarks of biological motion processing is how rapidly the coherent perception of human action arises. Johansson's early work suggested that explicit detection and recognition of a variety of individual actions could take place in as little as 200 ms (Johansson, 1973), with very similar estimates (232 ms) arising from later visual search studies involving direction discrimination (Cavanagh et al., 2001). While the estimation of speed from point-light walkers is likely to take additional time – given the need for some level of temporal integration – such processes could still be highly efficient (Festa & Welch, 1997; McKee & Welch, 1985; Watamaniuk & Duchon, 1992), and certainly achieved within 500 ms (Mather & Parsons, 2018).

In Experiment 1, the dynamic arrays of walkers were shown for 3 seconds. Observers could thus have serially sampled up to 6 walkers, assuming the liberal estimate of 500 ms per

walker just mentioned. Such serial sampling strategies have been proposed (Myczek & Simons, 2008) as alternatives to the idea that ensemble processes occur automatically and in parallel across the visual field (Chong & Treisman, 2003; Treisman, 2006).

In Experiment 2 we attempted to address this issue by systematically reducing stimulus viewing duration to a degree that any explicit calculation of the average crowd speed would become highly unlikely. Specifically, beginning with displays shown for 2500 ms, we gradually reduced viewing time by 500 ms across five consecutive blocks of trials until observers only had 500 ms to view the crowds. If reduced viewing time leads to a consistent reduction in estimation ability, this would suggest that observers need to engage in serial sub-sampling in order to judge crowd speed.

Additionally, the efficiency of ensemble perception can be estimated by various simulation techniques (Baek & Chong, 2020; Dakin, 2001; Haberman & Whitney, 2007; Maule & Franklin, 2016; Solomon et al., 2011). Following these studies, we conducted a series of simulations to estimate the number of PLWs integrated during ensemble processing of crowd speed as a function of viewing duration.

Method

Participants

Sixteen observers (M_{age} = 22.9 years, SD_{age} = 3.1; 5 female; 14 right-handed) were recruited from the research community at the University of Malta. Sample size was determined as in Experiment 1. Five observers had already taken part in the first study while the remaining had no experience in psychophysical experiments. All observers had normal or corrected-to-normal vision with no colour deficiency. Observers gave written consent before the experiment and were monetarily compensated. The experiment lasted approximately 45-60 minutes. All methods and procedures conformed to the Ethics and Data Protection Guidelines of the University of Malta.

Visual Stimuli, Task, and Procedure

As we were interested in assessing biological motion processing under optimal conditions -- where both local and global information is available -- Experiment 2 only made use of an intact-PLW crowd while keeping the probe array, response array, and response method the same as in Experiment 1. We also retained the same speed profile ranging from 40 steps/min to 150 steps/min with 10 steps/min increments (corresponding to 1-12 speed values)

As in Experiment 1, we initially presented the homogeneous condition as practice trials with a fixed trial duration of 3 seconds. In subsequent blocks involving the heterogeneous condition, we systematically reduced viewing duration as described in more detail shortly. In the heterogeneous condition, we limited crowd speed between 3 and 10 speed values with a constant crowd variability at three speed units. On a given trial, average crowd speed was first selected randomly and uniformly from the truncated speed range. Individual speeds of a given crowd were then selected randomly from a normal distribution with a previously selected mean and a standard deviation of 3 speed units. In contrast to Experiment 1, we ensured a balanced design in which participants viewed each crowd speed value for the same amount of trials. Thus, we applied a threshold of ±0.25 unit around the requested mean speed, together with the existing threshold of ±0.5 unit around the requested standard deviation.

We gradually reduced viewing duration across five blocks of trials, each block consisted of 80 trials making up a total of 400 trials (5 blocks x 8 speeds x 10 repetitions). Participants viewed the crowd stimulus for 2500 ms in the first block. In each subsequent block, display duration was reduced 500 ms. Thus, in the last block of trials, the crowd stimulus was shown for only 500 ms. We did not counterbalance viewing durations because our pilot study showed that participants initially struggled to do the task at very short viewing duration. This could introduce potential confounding effects due to the order of viewing durations. However, our additional analysis showed that order effect was not present in our task (see details in Supplementary Materials).

Data Analysis

We first analysed the effect of stimulus duration using 5 (Duration) x 8 (Speed) ANOVAs. Separates ANOVAs were conducted for perceived crowd speed and response variability. We then determined whether observers were able to integrate more than one walker to compute the average crowd speed at 500 ms. Similar to Experiment 1, we conducted single-walker response simulation and compared that to response variability obtained at 500 ms. We applied Bonferroni correction for multiple testing and Greenhouse-Geisser's correction for sphericity violations where necessary. Finally, we carried out ideal-observer simulations to estimate the number of PLWs integrated at short and long viewing durations.

Data Availability

The data and analysis routines relating to this experiment can be accessed via the associated OSF page at https://osf.io/5j4qe/.

Results

Figure 6a shows perceived crowd speed as a function of actual crowd speed and viewing durations. As in Experiment 1, the overall pattern of responses suggests that observers were able to accurately estimate the true mean speed. There again appears to be a tendency to overestimating slow crowds and underestimating fast crowds. Interestingly, overestimation of slow crowds appears to amplify as viewing duration reduces, a tendency we return to in the General Discussion.

The 5 (Duration) x 8 (Speed) ANOVA showed a significant main effect of Speed, $F(1.68, 25.24) = 129.40$, MSE = 5.61, $p < 0.001$, $\eta^2 = 0.90$. All post-hoc pairwise comparisons for crowd speed were significant, confirming that observers could accurately discriminate different crowd speeds (all ts > 4.47 and all ps < 0.05). The main effect of Duration was also significant, $F(4, 60) = 6.93$, MSE = 0.85, $p < 0.001$, $\eta^2 = 0.32$. Post-hoc tests showed that performance at 500 ms was significantly different from 2000 ms and 2500 ms (ts > 3.60 and ps < 0.05), and performance at 1000 ms was significantly different from 2000 ms and 2500 ms (ts > 3.56 and ps < 0.05). There was also a significant Duration x Speed interaction, $F(28, 420) = 1.56$, MSE = 0.30, $p = 0.037$, $\eta^2 = 0.09$, indicating that reduced viewing time led to decreased estimation accuracy but the effect varied according to stimulus speed.

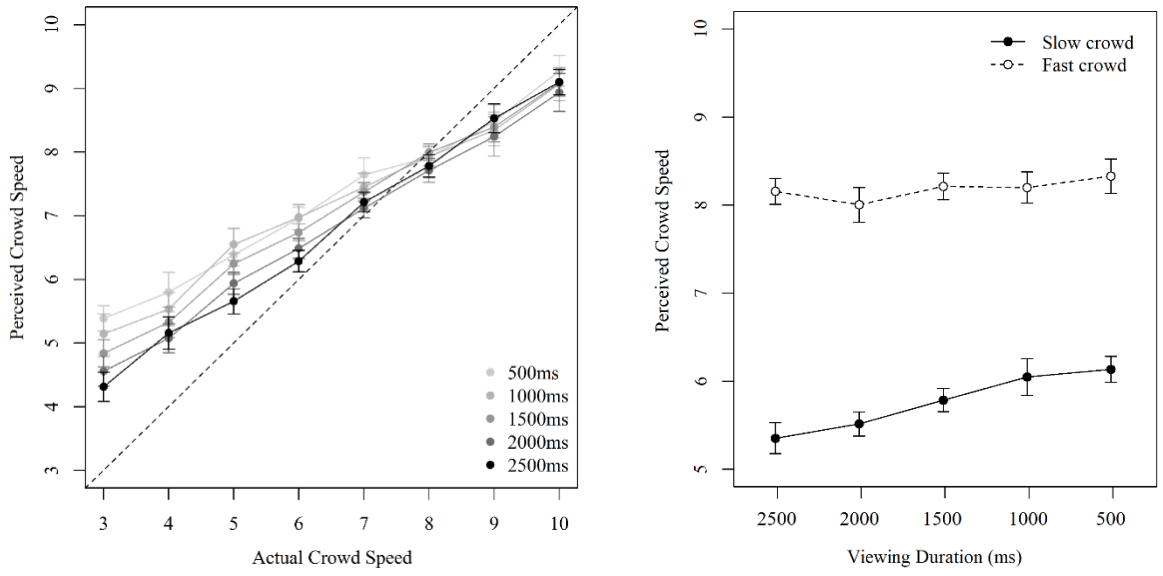

Figure 6. A) Perceived crowd speed as a function of actual crowd speed and viewing durations. B) Perceived crowd speed as a function of viewing duration for slow and fast crowds. Error bars are ±1 standard error.

The nature of this interaction is illustrated in Figure 6b. We found the distinction between slow and fast crowds highlighted in Experiment 1 particularly useful. We divided

speed range as previously -- speeds 3-6 for slow crowds and 7-10 for fast crowds -- and conducted simple ANOVAs with viewing duration as the independent variable. These showed a significant linear effect of Duration for slow crowds, $F(1, 15) = 45.58$, MSE = 0.154, $p < 0.001$, $\eta^2 = 0.75$; but no significant linear effect of Duration for fast crowds, $F(1, 15) = 2.04$, MSE = 0.23, $p = 0.17$, $\eta^2 = 0.12$. Thus, reduced viewing duration only affected estimation accuracy for slow crowds.

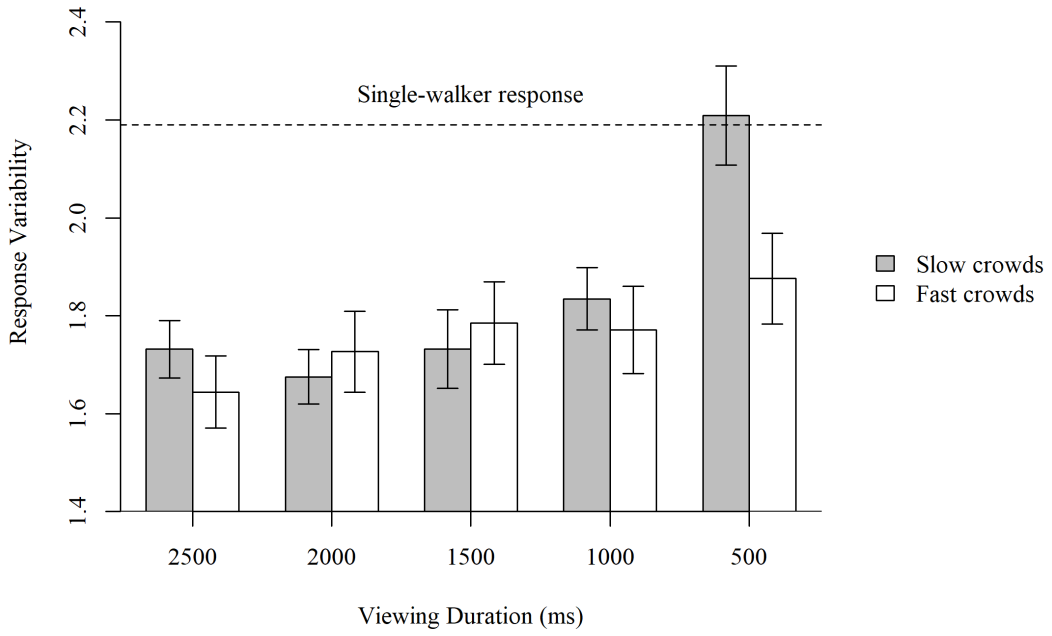

Figure 7. Response variability across viewing duration for grouped slow and fast crowds. Error bars are ±1 standard error.

In terms of response variability, the 5 (Duration) x 8 (Speed) ANOVA showed a significant main effect of Duration, $F(4, 60) = 7.19$, MSE = 0.33, $p < 0.001$, $\eta^2 = 0.32$. Post-hoc tests indicated that response variability at 500 ms was significantly higher than at other durations except for 1500 ms (1500 ms: $t = 2.99$, $p = 0.09$; other durations: all ts > 3.58, ps < 0.05). There was also a significant main effect of Speed, $F(7, 105) = 2.71$, MSE = 0.27, $p = 0.013$, $\eta^2 = 0.15$, and a significant Duration x Speed interaction, $F(28, 420) = 1.81$, MSE = 0.15, $p = 0.008$, $\eta^2 = 0.11$. However, none of the post-hoc pairwise comparisons survived Bonferroni correction (ts < 3.26, ps > 0.15), and there was no clear pattern across all speeds to help with interpreting the nature of the interaction (see Supplementary Materials for further information).

We thus grouped crowd speed into slow and fast as previously and ran additional post-hoc tests on this data to understand the main effect of crowd speed and the interaction effect. Figure 7 shows response variability across viewing duration for the grouped slow and fast crowds. Comparing the effect of crowd speeds at each duration, we found that response variability of slow crowds was only significantly higher than fast crowds at a viewing duration of 500 ms ($t = 4.11$, $p = 0.004$). The remaining pairwise comparisons were non-significant (all ts < 2.08, ps > 0.05; before Bonferroni correction). Thus, only the combined effect of slow speeds and 500 ms viewing duration had a significant influence on response variability.

Figure 7 also illustrates how observed precision differs from a single-walker response strategy that was simulated as in Experiment 1. In all cases except slow crowds viewed for 500 ms, observers appear to be basing their speed estimates on an average of more than one walker. To further explore the differential performance with a 500 ms viewing duration, Figure 8 expands the block data to show the entire speed range. This confirms that for slow crowds (speed 3-6), observers did not perform significantly better than the single random walker simulation (ts < 2.12; ps > 0.05), but for fast crowds (speed 7-10), they did (ts > 2.50; ps < 0.02). Thus, even with a very brief viewing duration, observers were still able to integrate more than one walker to estimate fast crowds, but this was not case for slow crowds.

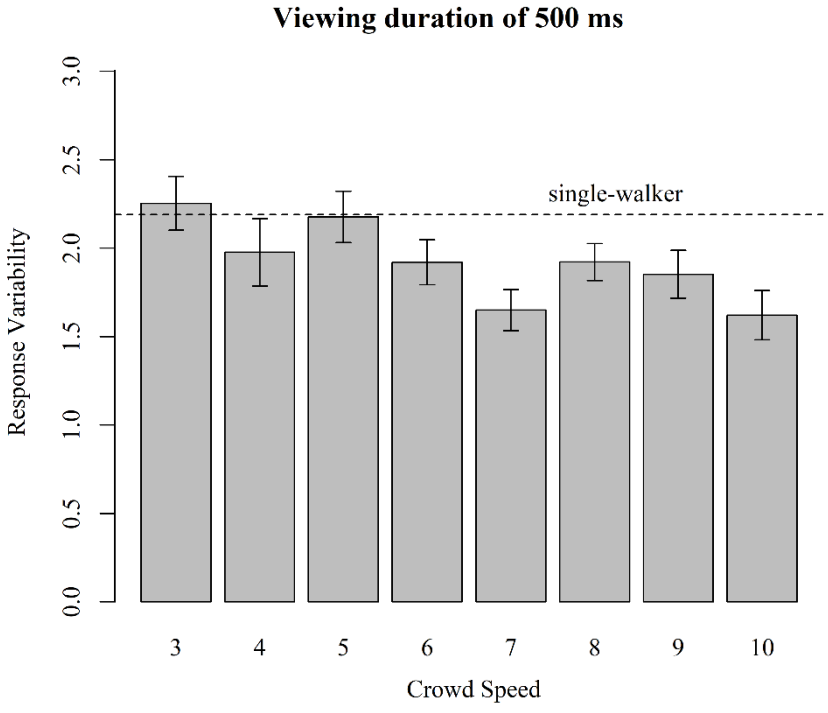

Figure 8. Response variability at 500 ms for each speed in comparisons to chance-level and single-walker based response simulations. Error bars are ±1 standard error.

Ensemble coding efficiency of crowd speed

Research in ensemble perception has frequently used ideal-observer models to investigate its possible underlying mechanism and integration efficiency (e.g., Allik et al., 2013; Baek & Chong, 2020; Myczek & Simons, 2008; Parkes et al., 2001; Solomon et al., 2011). While these modelling approaches vary in terms of decision structures, model components and/or assumptions, they all have highlighted the important role of internal noise to achieve better goodness-of-fit of a model. Internal noise reflects intrinsic random variations of the visual system that lead to different judgments of the same stimulus viewed on different occasions. In the case of ensemble perception, internal noise can occur before averaging takes place, i.e. noisy percept of individual crowd members and hence termed *early* noise. Internal noise can also occur post-averaging, i.e. noise percept of the average crowd characteristics and hence termed *late* noise. Here, we followed previous studies and incorporated internal noise component in our models to estimate the efficiency of ensemble perception of crowd speed.

Modelling approach

To further investigate the processing efficiency of ensemble perception of crowd speed, we conducted simulations to estimate the number of PLWs participants integrated in different heterogeneous-crowd conditions – what we refer to as the effective sampling size (n_e). We simulated ideal observers which randomly sampled n PLWs to compute the average crowd speed while the remaining unsampled PLWs would not contribute to the final crowd speed estimation (see Baek & Chong, 2020 for an alternative approach). We used participants' response variability obtained from the homogeneous-crowd condition to estimate the overall amount of internal noise which was then used to construct three different ideal-observer models: *early* noise only, *early* + *late* noise, and *late* noise only. Finally, we intersected the model prediction curves and actual response variability in the heterogeneous-crowd condition to derive the effective sampling size n_e.

Simulation procedure

For each participant, the simulation procedure had three phases: crowd generation, sampling, and averaging. Figure 9 illustrates these phases and how internal noise contributes to the speed estimation for the different models. In the first phase, we generated 100 samples (each sample has 120 trials, making a total of 12,000 trials for each participant) using the exact same methodology used to generate heterogeneous crowd arrays in the empirical task. We then simulated the response of each trial with n equal to 1, 2, and so on, until n = 12. The simulated

trial-by-trial responses were used to generate a simulated response-variability curve (i.e. response variability as a function of sampling size, *n*).

For the *early* noise model, we applied internal noise during the sampling phase, that is, noise was added to each of the *n* sampled PLWs before averaging their speeds. In the *late* noise model, we applied internal noise during the averaging phase. That is, we first sampled *n* PLWs and calculated their average speed. We then applied internal noise to the averaged speed.

For the *early* + *late* noise model, because our tasks were not designed to quantify *early* and *late* noise separately, we relied on previous models to estimate the amount of each type of noise. Specifically, our model resembled the no-attention with *late* noise model of Baek and Chong (2020), which specifies how *early* noise (σ_e) and *late* noise (σ_l) relate to the total amount of internal noise (ζ) as follows:

$$\zeta = \sqrt{\frac{\sigma_e^2}{N} + \sigma_l^2}$$

There were several assumptions we made for our simulations. First, we assumed that internal noise had a Gaussian distribution. Second, we assumed that internal noise was invariant to actual stimulus speed. Our last two assumptions were specific to the equation above for the *early* + *late* noise model: we assumed $N = 12$, i.e. all set members of the homogeneous crowds, and $\sigma_e \sim 2\sigma_l$. The latter assumption roughly matched with the best fit result of the no-attention with *late* noise model (Table 4, Baek & Chong, 2020, p. 77). These assumptions were arbitrary and changing them would lead to slightly different estimates. Nevertheless, they are sufficient for our modelling purposes.

Finally, we intersected each participant's actual response variability in a given heterogeneous condition with the ideal observer's response-variability curve for that condition. This intersection indicated the estimated value of n_e – the effective sampling size needed to achieve the observed performance. We conducted separate simulations for fast and slow crowds viewed at 500 ms and 2500 ms. The distinction between fast and slow crowds came from the analysis of empirical results as reported previously.

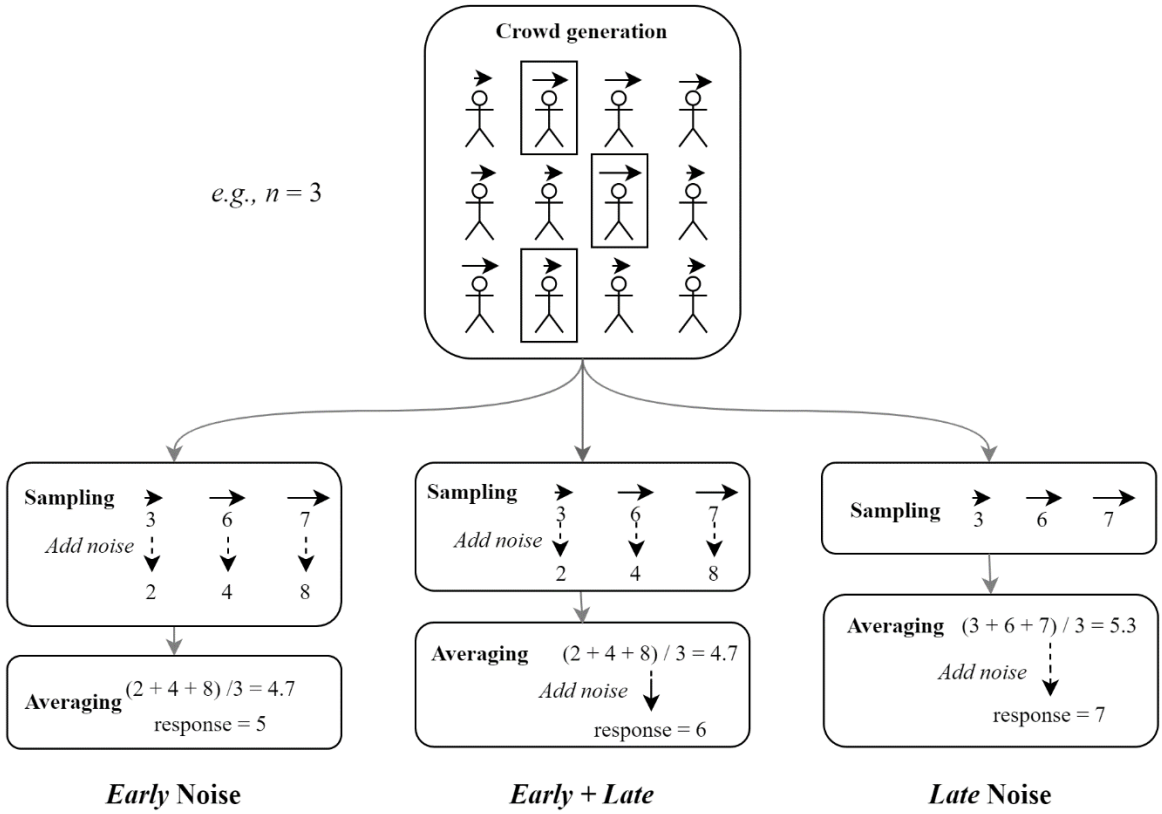

Figure 9. Simulation procedure for *early*, *early* + *late* and *late* noise models to estimate the number of PLWs integrated in ensemble processing of crowd speed.

Simulation results and discussion

Figure 10 shows an example of the response-variability curves of the three models using data of Participant 1. The *early* noise model predicted that response variability sharply reduced as the number of integrated PLWs increased. This was due to noise cancelation which occurred during the averaging phase. As sampling size increases, the *early* noise model's prediction of response variability quickly exceeds the baseline condition of homogeneous crowds. The *early* noise model thus imposed a very conservative limit on the estimates of sampling size.

On the other hand, the *late* noise model showed a gradual reduction in response variability because noise was added after averaging, and hence was unaffected by sampling size. The addition of *late* noise reflects actual human performance, that is, mean estimation improves and then plateaus as set size continues to increase. Thus, *late* noise is considered an important factor limiting the processing capacity of ensemble perception (Baek & Chong, 2020; Parkes et al., 2001; Solomon et al., 2011). The *late* noise model, however, may estimate n_e too liberally, that is, n_e can reach full set size as performance in the heterogeneous condition approaches performance in the homogeneous condition.

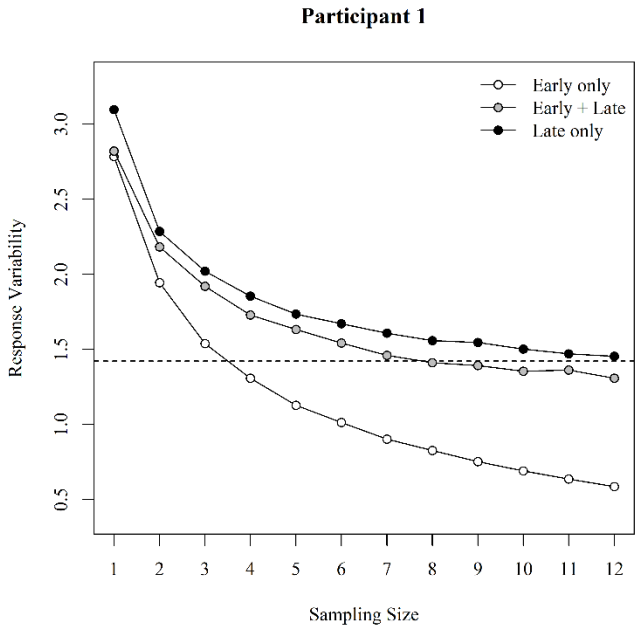

Figure 10. Response-variability curve for the *early*, *early* + *late* and *late* noise models using the internal noise of Participant 1 to simulate the full range of crowd speed. Horizontal dashed line indicates the overall amount of internal noise obtained from the homogeneous condition.

Table 1 presents the estimated mean effective sampling size for all three models as a function viewing duration and crowd speed. However, as illustrated in Figure 10, the *early* + *late* noise model appears to be the most appropriate for our modelling purpose. It retained the plateaued performance with increasing sampling size limited by *late* noise and provided more conservative estimates of n_e compared to the *late* noise only model. We thus based our conclusions on the estimates provided by the *early* + *late* noise models.

Table 1: Mean (standard error) of the number of integrated PLWs (effective sampling size) estimated by the early and late noise models in 2 (Duration) x 2 (Crowd Speed) conditions.

	500 ms		**2500 ms**	
	Slow	**Fast**	**Slow**	**Fast**
***Early* only**	1.85 (0.10)	2.41 (0.13)	2.53 (0.12)	2.79 (0.16)
Early* + *Late	2.37 (0.16)	3.32 (0.26)	3.70 (0.30)	4.82 (0.57)
***Late* only**	2.21 (0.17)	4.23 (0.67)	3.81 (0.34)	5.41 (0.41)

Figure 11 shows the simulated response-variability curve as a function of crowd speed for the *early* + *late* noise model. At 500 ms, the model estimated that participants integrated around 2-3 PLWs (M = 2.37, SE = 0.16) to compute the average speed of slow crowds (3-6 speed units). For fast crowds (7-10 speed units), the number of integrated PLWs at 500 ms was approximately 3-4 (M = 3.32, SE = 0.26). At our longest viewing duration (2500 ms), the model estimated approximately 3-4 PLWs for slow crowds (M = 3.70, SE = 0.30), and 4-5 PLWs for fast crowds (M = 4.82, SE = 0.57), suggesting that there may be an upper limit to the number of walkers that can be integrated.

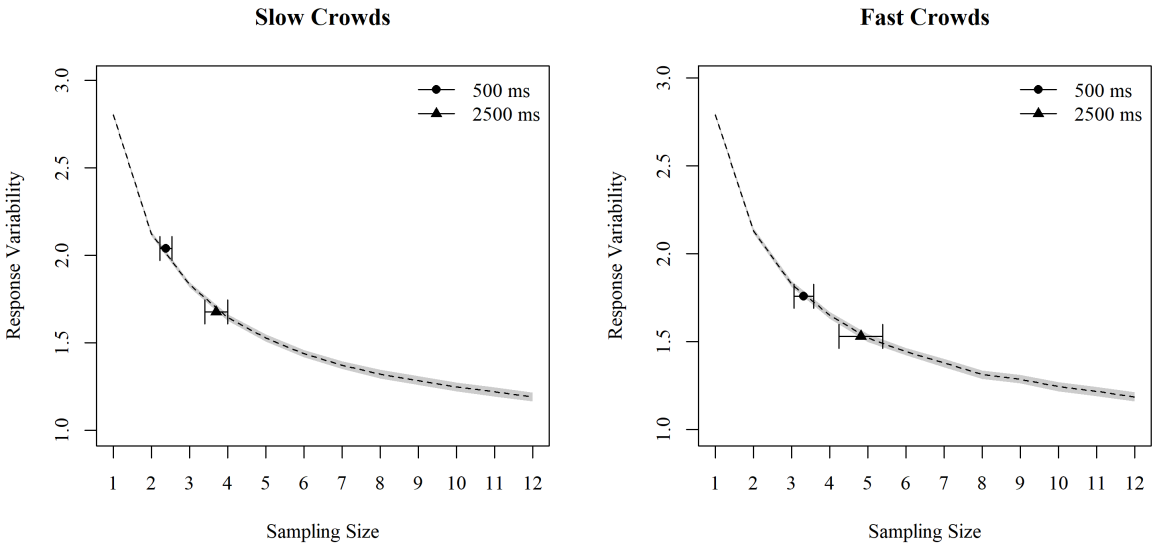

Figure 11. The average number of PWLs integrated to ensemble perception of fast and slow crowds viewed at 500 ms and 2500 ms, simulated based on the *early* + *late* noise models. Error bars and shading areas are ±1 standard error.

As mentioned previously, our modelling approach assumed a specific mechanism of ensemble perception and hence the value of n_e carried a specific meaning. For example, an estimated n_e = 4 would indicate that participants could utilise the information from approximately four walkers to compute the average and ignore the remaining eight walkers in the crowd. However, this is not necessarily the actual mechanism of ensemble perception. For example, Baek and Chong (2020) proposed an alternative model assuming a distributed attention component that accounted well for performance in their mean size estimation tasks. Under this model's assumptions, all set members can contribute to mean estimation and the amount of contribution from each set member is mediated by attentional allocation and set size. Thus, the value of n_e within this view can also be interpreted as the proportion of the total available information provided by the stimulus set. As these authors have pointed out, more

work is needed to differentiate between the two mechanisms. Therefore, our model estimates at best should be taken as an indication of integration efficiency without affirming the exact mechanism involved.

Discussion

In Experiment 2, we investigated the possibility the observers use a sub-sampling strategy when estimating crowd speed by gradually reducing stimulus exposure. Starting at 2500 ms in the first block, viewing duration was reduced by 500 ms across five blocks of trials. In the last block, observers had only 500 ms to estimate the average speed of a crowd. If performance systematically deteriorated across the full range of speeds, this would indicate that ensemble processing of crowd speed does not occur automatically, and observers could rely on serial subsampling strategies. Our results, however, showed that viewing durations affected ensemble perception of crowd speed differently for fast and slow speed ranges.

More specifically, when the crowd moved at fast speeds (> 70 steps/min), observers could estimate the average accurately and with high precision regardless of viewing duration. There was thus no evidence that participants used serial sub-sampling with crowds within this range. Furthermore, our observer models found that around 3-4 walkers were integrated in ensemble coding of crowd speed at 500 ms which is impressive because the participant would be able to make few eye movements during this time. Overall, the lack of dependence on viewing duration, accurate performance at very brief viewing durations, and a high degree of encoding efficiency suggests that speed perception of fast crowds can be automatic.

In contrast to fast crowds, we found that estimation of slow crowds suffered from reduced viewing duration. There appears to be a systematic deterioration in perceptual accuracy of slow crowds in which observers tended to overestimate the average crowd speed as the crowd moved more slowly. In terms of precision, observers' performance became significantly worse at 500 ms and their performance was not better than the single-walker response strategy. At these speeds, there is relatively little dynamic information available. This could lead to ensemble mechanisms assigning more weight to faster speeds or it is possible that participants resort to an explicit, serial strategy of some form. For example, walkers that present a greater proportion of the step-cycle could be given more weight in the averaging because their speed estimate is more precise. Paying more attention to the faster walkers in the array thus could be a viable sampling strategy which might also explain the tendency to overestimate speed with slow crowds. Of course, implicit averaging mechanisms may also be sensitive to variations in

reliability as a function of speed, assigning more weight to faster figures within the crowd when computing ensemble estimates (de Fockert & Marchant, 2008; Ernst & Banks, 2002; Haberman & Whitney, 2010; Kanaya et al., 2018). Such an idea clearly warrants further investigation.

However, as noted in the introduction to Experiment 2, the nature of our stimuli makes it impossible to rule out that the observed deterioration in performance was simply due to the very sparse speed information that would be available in 500 ms for slowly moving figures. This would be expected to impact both serial and parallel approaches to estimation.

General Discussion

In two experiments, we have demonstrated that the human visual system is sensitive to the average speed of a crowd. This ability was observed for crowds containing either locally scrambled or globally intact PLWs (Experiment 1) and persisted even when intact-PLW crowds were displayed for a very brief duration (Experiment 2). These findings extend the existing literature which demonstrates that human observers are capable of extracting summary statistics of crowd information, such as average gender and facial expression (Haberman & Whitney, 2007), family resemblance (de Fockert & Wolfenstein, 2009), gaze direction (Sweeny & Whitney, 2014), and average crowd heading direction (Sweeny et al., 2012, 2013). As the first study to investigate averaging in the context of crowd speed, this work contributes both to the literature on biological motion processing (Mather & Parsons, 2018; Ueda et al., 2018) and the literature on ensemble perception (Whitney et al., 2014; Whitney & Yamanashi Leib, 2018), supporting the notion that the computation of summary statistics during everyday perception is robust and ubiquitous.

Contributions of global and local motion to ensemble speed estimation

We found that ensemble perception of crowd speed involved both local and global motion mechanisms. In Experiment 1, observers could estimate the average speed of a crowd made of scrambled PLWs. Crowd speed estimation can thus rely on local motion alone, possibly via compulsory information pooling mechanisms, as suggested by research using random dot kinematograms (Watamaniuk & Duchon, 1992). However, integrating local dots into dynamic human figures was beneficial to ensemble speed perception. This is consistent both with the general biological motion literature, where global advantages are often found (Bertenthal & Pinto, 1994; Bülthoff et al., 1998; Pavlova & Sokolov, 2000; Thornton & Vuong,

2004) and the specific finding that a global percept can improve speed discrimination performance for a single PLW (Ueda et al., 2018).

In the current context, it seems likely the availability of a global configuration might aid ensemble processing by providing more detailed information about the occurrence of coordinated "steps" (Neri et al., 1998) and/or by reducing spatial uncertainty as to the location of the most informative ankle/wrist dots (Cai et al., 2011; Mather et al., 1992). As biological motion processing is known to involve both bottom-up (Mather et al., 1992; Thornton et al., 2002; Troje & Westhoff, 2006) and top-down (Bülthoff et al., 1998; Cavanagh et al., 2001; Thornton et al., 2002) mechanisms (see Thornton, 2012 for a review), the current global advantage could also reflect the engagement of stored speed templates specific to human locomotion patterns, although our current data do not directly speak to this point.

Parallel or Serial Processing?

Of theoretical interest, there has been an ongoing debate as to whether ensemble perception of high-level representations emerges from parallel or serial processing. Treisman (2006) proposed that statistical moments are automatically computed under distributed attention via a parallel processing mechanism, integrating information across multiple objects. On the other hand, Myczek and Simons (2008) showed that simulations of various subsampling strategies using one to two set members could sufficiently explain observed performance, suggesting that ensemble perception can be produced cognitively via a serial inspection of a few set members and need not be automatic or unconstrained by limited attentional capacity. Others have argued for a middle ground account in which the visual system does indeed engage in ensemble processing, which has a limited capacity (Allik et al., 2013; de Fockert & Marchant, 2008; Haberman & Whitney, 2009; Maule & Franklin, 2016).

Here, for faster crowds, we found supporting evidence for the parallel processing account. Estimation precision and accuracy of fast intact-PLW crowds were not affected by stimulus exposure and simulations suggest that approximately 3-4 walkers were integrated within 500 ms. In the context of a dynamically evolving pattern, the ability to integrate 3-4 items within 500 ms argues strongly against the idea of explicit serial sub-sampling of individual items. For example, simple recognition of a single point-light action or visual search for an odd-one-out PLW based on direction (Cavanagh et al., 2001) require at least 250 ms, and speed perception is generally thought to take longer than direction discrimination (De

Bruyn & Orban, 1988). We note, however, that while we have shown that crowd speed perception was more reliable in the presence of global configurations (Experiment 1), our current design does not allow us to uniquely separate the contribution of local and global motion. Thus, the precise nature of the representations involved during such parallel processing of crowd speed estimation requires further investigation.

A different pattern was observed for slow moving crowds. Reducing stimulus duration led to lower estimation precision and accuracy for intact-PLW crowds. Furthermore, estimation of slow crowds at the briefest stimulus exposure was no better than predicted by a single-walker response strategy. These patterns suggest that when faced with limited information on which to base speed estimates (i.e., the slowest figure would only complete 1/3 of a step in 500 ms), participants may switch strategy and attempt to explicitly sub-sample the displays. However, given the results from faster speeds just noted, a more parsimonious explanation for the observed performance might be that the same mechanisms/strategies are applied to all types of crowds, but that weak low-level speed information contained in slow moving stimuli lead to much noisier speed estimates. In the current data, we are unable to distinguish between these two explanations for performance with slow crowds.

Overall, then -- at least when sufficient speed info is available -- it appears that participants are able to extract summary statistics from dynamic crowds in parallel. While impressive, a maximum number of 4-5 integrated items still suggests that ensemble perception is unable to utilise all information available (i.e., 12 figures) and/or that the square root of set size is some fixed limit for ensemble processing (Dakin, 2001). Whether additional information is unregistered or becomes compressed in some way is outside of the scope of the current study, nevertheless, this will be an interesting topic to be explored in future studies (see Alvarez, 2011 for further discussion).

Implicit versus explicit averaging?

We should note that while our results suggest that ensemble speed can sometimes be estimated in parallel, in the current task, participants were directed to explicitly pay attention to the entire array when making their judgements. A stronger test for such computations being truly "automatic" would be a situation where ensemble estimates could be shown to exist and to affect behaviour when attention is directed elsewhere (e.g., Fischer & Whitney, 2011; Oriet & Hozempa, 2016). Such implicit or incidental effects have been shown previously with

biological motion stimuli, at least for simple displays and direction judgements (e.g., Bosbach et al., 2004; Thornton & Vuong, 2004). For example, Thornton & Vuong (2004) adapted the classic Eriksen & Eriksen (1974) flanker task to show that the facing direction of to-be-ignored walkers influenced the reaction time to a central target figure. It would be interesting to explore this finding in the context of speed judgements and to extend such tasks to situations where flanking stimuli exert an influence as an ensemble, rather than individually. We note, however, that initial attempts to extend the flanker paradigm to include biological motion crowds with constant walking speeds have so far given rise to inconclusive results (Thornton et al., 2019).

Of course, if complex displays such as the ones used in the current study are divided into target and flanking stimuli, this may also raise the question of crowding, that is, difficulties in explicitly recognising or discriminating objects when they are presented in clutter (Bouma, 1970; Flom et al., 1963; Herzog et al., 2015; Levi, 2008; Manassi & Whitney, 2018; for a review, see Pelli & Tillman, 2008; Whitney & Levi, 2011). While a detailed discussion of the relationship between ensemble processes and crowding mechanisms is beyond the scope of the current paper (see Whitney & Yamanashi Leib, 2018), we mention the phenomenon here as there is evidence that biological motion displays are sensitive to visual clutter (Ikeda et al., 2013; Ikeda & Watanabe, 2016). It would certainly be interesting to explore whether crowding effects extend to the domain of speed judgements and more generally whether they might exert influences in complex arrays, such as those used here.

Overall though -- and we can only speculate – it seems unlikely that crowding would affect performance in the current task. First, the task instructions emphasise attention to the entire set rather than individual items. Second, if serial sampling of individual targets does occur under some conditions (e.g., slow motion with long display durations), such processing is likely to be accompanied by eye movements and centrally fixated targets are not typically thought to be affected by crowding.

More generally, we share the view expressed by Manassi & Whitney (2018), that crowding likely impairs explicit access to individual target information at various levels of the visual hierarchy, and does not degrade the object representations themselves, leaving them free to influence other aspects of behaviour. Depending on task instructions, then, the same display can give rise to both crowding and ensemble effects, with the two processes not necessarily interacting (Bulakowski et al., 2011).

The idea that there is a dissociation between limits on explicit reports and the influence of implicit representations resonates well with other areas of visual processing, for example implicit change detection (Fernandez-Duque & Thornton, 2000, 2003; Laloyaux et al., 2006), and may even relate directly to differences seen in the role of attention in explicit (Cavanagh et al., 2001; Thornton et al., 2002) versus implicit (Bosbach et al., 2004; Thornton & Vuong, 2004) processing of biological motion (Thompson & Parasuraman, 2012; see Thornton, 2012 for further discussion). In our everyday lives, it is likely that we routinely engage in both implicit and explicit processing of biological motion. One idea that may warrant further investigation is that the former is tuned by default to extract information from multiple sources – allowing us to quickly extract the "gist" of the behaviour of those around us– while the latter is tuned for the more detailed examination of a single actor (Rensink, 2000; Thornton et al., 2002).

General influence of speed range on ensemble perception

In addition to the influence on estimation strategies just mentioned, the speed range of a crowd also gave rise to other interesting patterns. For example, the perception of slower crowds was more prone to either spatial disruption (Experiment 1) or temporal manipulation (Experiment 2), compared to the perception of faster crowds. Furthermore, in Experiment 1, observers appeared to perceive intact-PLW crowds as slower than scrambled-PLW crowds. This pattern is consistent with the '*global slowdown effect*' reported in perceptual grouping research (Kohler et al., 2014) as well as speed perception research using single PLWs (Ueda et al., 2018). The '*global slowdown effect*' actually improves perceptual accuracy of crowd speed but only for slow crowds, suggesting that ensemble representations of fast crowds are more accurate and less susceptible to added local noise (i.e., scrambled PLWs are noisier than intact PLWs). In Experiment 2, reducing viewing duration led to more perceptual bias of crowd speed but again only for slow crowds. In addition, the efficiency of information integration was worst for slow crowds exposed for very brief duration. Together, these results suggest that observers were better at speed estimation of fast crowds than slow crowds.

The performance discrepancy between fast and slow crowds could be due to discrimination thresholds for retinal velocity of low-level motion. For slow PLWs, the ankle dots moved at approximately 2.5 – 3.8 arc/sec, and these are the fastest-moving dots in the figure. Discrimination thresholds for this speed range is more than 7% and up to 20% (De Bruyn & Orban, 1988). Velocity differences between two PLWs in our slow speed range were

between 11 and 17%, which might be lower than the limit of speed discrimination. Observers thus might have found it difficult to choose among slow speed options. On the other hand, the ankle dots of fast PLWs moved at approximately 4.2 – 5.4 arc/sec, and observers could discriminate speed change at a minimum 7% of reference speed (De Bruyn & Orban, 1988). Velocity differences between two fast PLWs in our task ranged between 8% and 10% and are slightly above discrimination threshold, which could explain why observers were more reliable in estimating fast crowds.

A useful approach to further explore the possible influence of retinal velocity in this context would be to vary the size/apparent distance of the individual walkers within the probe arrays. The retinal speed of a walking figure halves with each doubling of viewing distance. However, our perception of "actual" speed appears to be invariant to such changes. That is, people do not appear to move more slowly as they are viewed from further away (Mather et al., 2017; Mather & Parsons, 2018). More generally, our perception of action has been shown to operate very effectively even at quite extreme apparent distances (Thornton et al., 2014). In the current displays, all figures subtended 4.7° visual angle which -- assuming a standard physical height of 1.75 m -- is equivalent to looking at the crowd from approximately 20 meters away. Our displays could easily be modified so that the height of individual figures within a probe array was randomly varied, consistent with the impression of a crowd dispersed in depth. This would introduce considerable additional variation in terms of retinal speeds, without affecting the true "actual" speed of any given figure and thus the physical average speed. If ensemble estimates for the speed of such depth-dispersed crowds were similar to those observed in the current experiments, this would argue against a strong dependence on retinal velocity.

Aside from the influence of retinal velocity, better estimation accuracy of fast crowds could be due to our use of a fixed speed increment across speed range. Jacobs and Shiffrar (2005) reported that observers were worse at discriminating fast PLWs than slow PLWs, suggesting that speed perception of biological motion conforms to Weber's law. If this is the case, when comparing a pair of PLWs which have the same difference in speed, a pair of slow PLWs would appear more distinct than a pair of two fast PLWs, making a slow crowd appear more heterogeneous than a fast crowd despite that objectively both crowds have the same speed variance. This possibility would explain why observers are better at estimating the average speed of faster crowds because faster crowds would induce less perceptual crowd

variance. However, these explanations need to be confirmed by further psychophysical studies of speed perception using biological motion.

In addition to speed discrimination functions, the minimum time required to estimate absolute walking speed for single PLWs does not appear to have been previously reported, nor whether such estimates would be critically dependent on speed range. Based on the available data, we can presently conclude that observers were better at estimating the speed of fast-moving crowds than slow-moving crowds. Such a conclusion could carry some ecological validity, as a fast-moving crowd might signal more potential disturbances and require immediate action compared to a slow-moving crowd.

The effect of heading direction on crowd speed estimation

As noted in Experiment 1, an interesting avenue for future research would be to directly test the effect of heading direction on estimates of crowd speed. To explore this relationship, we binned data into two broad heading directions (sideways- and forward/backward-facing, see Supplementary Material). There were several preliminary findings of interest. First, we found that side views of crowds enhanced speed estimation precision relative to front/back view, although such effects were observed only in Experiment 2. Second, in terms of perceived crowd speed, observers consistently judged sideways-facing crowds faster than forward/backward-facing crowds in all conditions except the homogeneous condition of Experiment 2. Finally, the tendency to overestimate the speed of sideways-facing crowds compared to forward/backward-facing crowds was largest at 500 ms.

Salience-based encoding mechanism

In both experiments, we noted a tendency to overestimate slow crowds and underestimate fast crowds, particularly in the case of heterogeneous crowds. These tendencies could reflect response factors such as a range effect and/or a regression effect related to magnitude matching response methods (Crawford et al., 2019; Petzschner et al., 2015). Alternatively, Kanaya et al. (2018) have reported that mean perception can be preferentially weighted towards extreme salient set items – a perceptual bias which the authors termed the amplification effect. While the current experiments were not designed to fully address this issue, our additional amplification analysis offered reasons to suspect that salience-based encoding mechanism might also occur in our task (see Supplementary Materials).

Specifically, for slow crowds in both experiments, we found that observers perceived crowd speed accurately when all figures moved at the same speed, suggesting little contribution of decision factors relating to the current response method. When individual speeds varied, observers consistently overestimated crowd speed. Such results indicated that speed perception of slow crowds could be amplified by the existence of salient fast walkers. In addition, Kanaya et al. (2018) have also highlighted that amplification effect is asymmetrical and mean estimation was skewed towards items larger in size or flickering faster. This would be congruent with our findings that no clear indications of amplification effect were present in the case of fast crowds.

Biological motion as a useful test case for ensemble perception

Previous research has found that ensemble perception operates within discrete levels of the visual processing hierarchy (e.g., Ariely, 2001; Chong & Treisman, 2003; Dakin & Watt, 1997; Maule et al., 2014; Sweeny et al., 2013; Sweeny & Whitney, 2014). While it has been suggested that there is little correlation between low-level and high-level ensemble representations (Haberman et al., 2015; but see Florey et al., 2016), this conclusion is based on studies that have examined very different types of stimuli across the levels. As already noted, biological motion is thought to engage processing mechanisms at various levels of the hierarchy and can be analysed in terms of both low-level motion features and high-level, conceptual content.

Using biological motion, previous studies have shown that perception of crowd orientation requires intact high-level structure (Sweeny et al., 2012; Sweeny & Whitney, 2014). Perception of crowd speed, on the other hand, appears to involve multiple perceptual stages that interact. For example, high-level representations can exert positive influences on ensemble processing of low-level inputs by improving mean estimation reliability and accuracy. At the same time, the efficiency of ensemble processing of high-level representations seems to be constrained by the strength of low-level motion inputs, as in the case of differential performance between fast and slow crowds under brief exposure in Experiment 2. Although we can only speculate, one possibility is that ensemble perception of crowd speed involves a dynamic reciprocal relationship between low- and high-level representations. Such ideas have increasingly gained attention in object and scene perception (Groen et al., 2017; Hochstein & Ahissar, 2002) as well as contemporary deep neural networks mimicking human visual perception (Kriegeskorte, 2015). We suggest that biological motion might prove an interesting

class of stimuli with which to further explore how ensemble perception fits into such interactive frameworks.

Conclusions

In conclusion, we have demonstrated that the average speed of a walking crowd can be extracted accurately and efficiently within a brief glance. Ensemble perception of crowd speed may rely both on compulsory information pooling of low-level motion signals, and global integration of high-level representations. Performance is affected by the absolute speed of the crowd, with better estimation for fast speeds, where efficiency approaches the square root of the crowd size. Biological motion perception may be a useful model for further ensemble perception studies, as it may engage ensemble coding mechanisms at multiple levels within the visual hierarchy. The current findings contribute to a growing literature showing that human observers are sensitive to crowd characteristics which in turn inform and facilitate social behaviours in public social gathering contexts.

Open Practices Statement

All data and analysis routines relating to this work can be accessed via the associated OSF page at https://osf.io/5j4qe/.

These experiments were not preregistered.

Acknowledgements

The work of TTNN and IMT was supported by research funds from the University of Malta. The authors are very grateful to Katsumi Watanabe and an additional anonymous reviewer for their very useful comments and suggestions while preparing this manuscript.

References

Allik, J., Toom, M., Raidvee, A., Averin, K., & Kreegipuu, K. (2013). An almost general theory of mean size perception. *Vision Research*, *83*, 25–39. https://doi.org/10.1016/j.visres.2013.02.018

Alvarez, G. A. (2011). Representing multiple objects as an ensemble enhances visual cognition. *Trends in Cognitive Sciences*, *15*(3), 122–131. https://doi.org/10.1016/j.tics.2011.01.003

Alvarez, G. A., & Oliva, A. (2008). The representation of simple ensemble visual features outside the focus of attention. *Psychological Science*, *19*(4), 392–398. https://doi.org/10.1111/j.1467-9280.2008.02098.x

Ariely, D. (2001). Seeing sets: Representation by statistical properties. *Psychological Science*, *12*(2), 157–162. https://doi.org/10.1111/1467-9280.00327

Atchley, P., & Andersen, G. J. (1995). Discrimination of speed distributions: Sensitivity to statistical properties. *Vision Research*, *35*(22), 3131–3144. https://doi.org/10.1016/0042-6989(95)00057-7

Baek, J., & Chong, S. C. (2020). Distributed attention model of perceptual averaging. *Attention, Perception, & Psychophysics*, *82*, 63–79. https://doi.org/10.3758/s13414-019-01827-z

Bauer, B. (2009). Does Stevens's Power Law for Brightness Extend to Perceptual Brightness Averaging? *The Psychological Record*, *59*(2), 171–185. https://doi.org/10.1007/BF03395657

Bertenthal, B. I., & Pinto, J. (1994). Global processing of biological motions. *Psychological Science*, *5*, 221–225.

Blake, R., & Shiffrar, M. (2007). Perception of Human Motion. *Annual Review of Psychology*, *58*, 47–73. https://doi.org/10.1146/annurev.psych.57.102904.190152

Boker, S. M., Cohn, J. F., Theobald, B. J., Matthews, I., Mangini, M., Spies, J. R., Ambadar, Z., & Brick, T. R. (2011). Something in the way we move: Motion dynamics, not perceived sex, influence head movements in conversation. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(3), 874–891. https://doi.org/10.1037/a0021928

Bolling, D. Z., Pelphrey, K. A., & Kaiser, M. D. (2013). Social inclusion enhances biological motion processing: A functional near-infrared spectroscopy study. *Brain Topography*, *26*(2), 315–325. https://doi.org/10.1007/s10548-012-0253-y

Bosbach, S., Prinz, W., & Kerzel, D. (2004). A simon effect with stationary moving stimuli. *Journal of Experimental Psychology: Human Perception and Performance*, *30*(1), 39–55. https://doi.org/10.1037/0096-1523.30.1.39

Bouma, H. (1970). Interaction Effects in Parafoveal Letter Recognition. *Nature*, *226*(5241), 177–178. https://doi.org/10.1038/226177a0

Brady, T. F., Shafer-Skelton, A., & Alvarez, G. A. (2017). Global ensemble texture representations are critical to rapid scene perception. *Journal of Experimental Psychology. Human Perception and Performance*, *43*(6), 1160–1176. https://doi.org/10.1037/xhp0000399

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, *10*(4), 433–436.

Bulakowski, P. F., Post, R. B., & Whitney, D. (2011). Reexamining the possible benefits of visual crowding: Dissociating crowding from ensemble percepts. *Attention, Perception, & Psychophysics*, *73*(4), 1003–1009. https://doi.org/10.3758/s13414-010-0086-2

Bülthoff, I., Bülthoff, H., & Sinha, P. (1998). Top-down influences on stereoscopic depth-perception. *Nature Neuroscience*, *1*(3), 254–257. https://doi.org/10.1038/699

Cai, P., Yang, X. Y., Chen, L., & Jiang, Y. (2011). Motion speed modulates walking direction discrimination: The role of the feet in biological motion perception. *Chinese Science Bulletin*, *56*(19), 2025–2030. https://doi.org/10.1007/s11434-011-4528-6

Cavanagh, P., Labianca, A. T., & Thornton, I. M. (2001). Attention-based visual routines: Sprites. *Cognition*, *80*(1–2), 47–60. https://doi.org/10.1016/s0010-0277(00)00153-0

Chang, D. H. F., & Troje, N. F. (2009). Acceleration carries the local inversion effect in biological motion perception. *Journal of Vision*, *9*(1), 19–19. https://doi.org/10.1167/9.1.19

Chong, S. C., & Treisman, A. (2003). Representation of statistical properties. *Vision Research*, *43*(4), 393–404. https://doi.org/10.1016/S0042-6989(02)00596-5

Cohen, C. J., Morelli, F., & Scott, K. A. (2008). *A Surveillance System for the Recognition of Intent within Individuals and Crowds*. 559–565. https://doi.org/10.1109/THS.2008.4534514

Crawford, L. E., Corbin, J. C., & Landy, D. (2019). Prior experience informs ensemble coding. *Psychonomic Bulletin & Review*, *26*, 993–1000. https://doi.org/10.3758/s134263-018-1542-6

Dakin, S. C. (2001). Information limit on the spatial integration of local orientation signals. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *18*(5), 1016–1026. https://doi.org/10.1364/josaa.18.001016

Dakin, S. C., & Watt, R. J. (1997). The computation of orientation statistics from visual texture. *Vision Research*, *37*(22), 3181–3192. https://doi.org/10.1016/s0042-6989(97)00133-8

De Bruyn, B., & Orban, G. A. (1988). Human velocity and direction discrimination measured with random dot patterns. *Vision Research*, *28*(12), 1323–1335. https://doi.org/10.1016/0042-6989(88)90064-8

de Fockert, J. W., & Marchant, A. P. (2008). Attention modulates set representation by statistical properties. *Perception & Psychophysics*, *70*(5), 789–794. https://doi.org/10.3758/pp.70.5.789

de Fockert, J. W., & Wolfenstein, C. (2009). Rapid extraction of mean identity from sets of faces. *Quarterly Journal of Experimental Psychology (2006)*, *62*(9), 1716–1722. https://doi.org/10.1080/17470210902811249

de la Rosa, S., Choudhery, R. N., Curio, C., Ullman, S., Assif, L., & Bülthoff, H. H. (2014). Visual categorization of social interactions. *Visual Cognition*, *22*(9–10), 1233–1271. https://doi.org/10.1080/13506285.2014.991368

Delorme, A., Rousselet, G. A., Macé, M. J.-M., & Fabre-Thorpe, M. (2004). Interaction of top-down and bottom-up processing in the fast visual analysis of natural scenes. *Cognitive Brain Research*, *19*(2), 103–113. https://doi.org/10.1016/j.cogbrainres.2003.11.01

Elias, E., Dyer, M., & Sweeny, T. D. (2017). Ensemble Perception of Dynamic Emotional Groups. *Psychological Science*, *28*(2), 193–203. https://doi.org/10.1177/0956797616678188

Eriksen, B. A., & Eriksen, C. W. (1974). Effects of noise letters upon the identification of a target letter in a nonsearch task. *Perception & Psychophysics*, *16*(1), 143–149. https://doi.org/10.3758/BF03203267

Ernst, M. O., & Banks, M. S. (2002). Humans integrate visual and haptic information in a statistically optimal fashion. *Nature*, *415*(6870), 429–433. https://doi.org/10.1038/415429a

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, *39*(2), 175–191. https://doi.org/10.3758/BF03193146

Fernandez-Duque, D., & Thornton, I. M. (2000). Change Detection Without Awareness: Do Explicit Reports Underestimate the Representation of Change in the Visual System? *Visual Cognition*, *7*(1–3), 323–344. https://doi.org/10.1080/135062800394838

Fernandez-Duque, D., & Thornton, I. M. (2003). Explicit mechanisms do not account for implicit localization and identification of change: An empirical reply to Mitroff et al. (2002). *Journal of Experimental Psychology: Human Perception and Performance*, *29*(5), 846–858. https://doi.org/10.1037/0096-1523.29.5.846

Festa, E. K., & Welch, L. (1997). Recruitment mechanisms in speed and fine-direction discrimination tasks. *Vision Research*, *37*(22), 3129–3143. https://doi.org/10.1016/S0042-6989(97)00118-1

Fischer, J., & Whitney, D. (2011). Object-level visual information gets through the bottleneck of crowding. *Journal of Neurophysiology*, *106*(3), 1389–1398. https://doi.org/10.1152/jn.00904.2010

Flom, M. C., Weymouth, F. W., & Kahneman, D. (1963). Visual Resolution and Contour Interaction*. *Journal of the Optical Society of America*, *53*(9), 1026. https://doi.org/10.1364/JOSA.53.001026

Florey, J., Clifford, C. W. G., Dakin, S., & Mareschal, I. (2016). Spatial limitations in averaging social cues. *Scientific Reports*, *6*(1), 32210. https://doi.org/10.1038/srep32210

Georgescu, A. L., Kuzmanovic, B., Santos, N. S., Tepest, R., Bente, G., Tittgemeyer, M., & Vogeley, K. (2014). Perceiving nonverbal behavior: Neural correlates of processing movement fluency and contingency in dyadic interactions: Perceiving Nonverbal Interactive Behavior. *Human Brain Mapping*, *35*(4), 1362–1378. https://doi.org/10.1002/hbm.22259

Giese, M. A., & Poggio, T. (2003). Neural mechanisms for the recognition of biological movements. *Nature Reviews Neuroscience*, *4*(3), 179–192. https://doi.org/10.1038/nrn1057

Grayson, B., & Stein, M. I. (1981). Attracting Assault: Victims' Nonverbal Cues. *Journal of Communication*, *31*(1), 68–75.

Groen, I. I. A., Silson, E. H., & Baker, C. I. (2017). Contributions of low- and high-level properties to neural processing of visual scenes in the human brain. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1714), 20160102. https://doi.org/10.1098/rstb.2016.0102

Haberman, J., Brady, T. F., & Alvarez, G. A. (2015). Individual differences in ensemble perception reveal multiple, independent levels of ensemble representation. *Journal of Experimental Psychology. General*, *144*(2), 432–446. https://doi.org/10.1037/xge0000053

Haberman, J., & Whitney, D. (2007). Rapid extraction of mean emotion and gender from sets of faces. *Current Biology*, *17*(17), 751–753. https://doi.org/10.1016/j.cub.2007.06.039

Haberman, J., & Whitney, D. (2009). Seeing the mean: Ensemble coding for sets of faces. *Journal of Experimental Psychology: Human Perception and Performance*, *35*(3), 718–734. https://doi.org/10.1037/a0013899

Haberman, J., & Whitney, D. (2010). The visual system discounts emotional deviants when extracting average expression. *Attention, Perception & Psychophysics*, *72*(7), 1825–1838. https://doi.org/10.3758/APP.72.7.1825

Herzog, M. H., Sayim, B., Chicherov, V., & Manassi, M. (2015). Crowding, grouping, and object recognition: A matter of appearance. *Journal of Vision*, *15*(6), 1–18. https://doi.org/10.1167/15.6.5

Hochstein, S., & Ahissar, M. (2002). View from the top: Hierarchies and reverse hierarchies in the visual system. *Neuron*, *36*(5), 791–804. https://doi.org/10.1016/s0896-6273(02)01091-7

Hu, Y., Baragchizadeh, A., & O'Toole, A. J. (2020). Integrating faces and bodies: Psychological and neural perspectives on whole person perception. *Neuroscience & Biobehavioral Reviews*, *112*, 472–486. https://doi.org/10.1016/j.neubiorev.2020.02.021

Ikeda, H., & Watanabe, K. (2016). Action Congruency Influences Crowding When Discriminating Biological Motion Direction. *Perception*, *45*(9), 1046–1059. https://doi.org/10.1177/0301006616651952

Ikeda, H., Watanabe, K., & Cavanagh, P. (2013). Crowding of biological motion stimuli. *Journal of Vision*, *13*(4), 1–6. https://doi.org/10.1167/13.4.20

Jacobs, A., & Shiffrar, M. (2005). Walking Perception by Walking Observers. *Journal of Experimental Psychology: Human Perception and Performance*, *31*(1), 157–169. https://doi.org/10.1037/0096-1523.31.1.157

Ji, L., & Pourtois, G. (2018). Capacity limitations to extract the mean emotion from multiple facial expressions depend on emotion variance. *Vision Research*, *145*, 39–48. https://doi.org/10.1016/j.visres.2018.03.007

Johansson, G. (1973). Visual perception of biological motion and a model for its analysis. *Perception & Psychophysics*, *14*(2), 201–211. https://doi.org/10. 3758/BF03212378

Johnson, K. L., & Shiffrar, M. (2013). *People watching: Social Perceptual, and Neurophysiological Studies of Body Perception*. Oxford University Press.

Kaiser, R., & Keller, P. E. (2011). Music's impact on the visual perception of emotional dyadic interactions. *Musicae Scientiae*, *15*(2), 270–287. https://doi.org/10.1177/1029864911401173

Kanaya, S., Hayashi, M. J., & Whitney, D. (2018). Exaggerated groups: Amplification in ensemble coding of temporal and spatial features. *Proceedings of the Royal Society B: Biological Sciences*, *285*(20172770), 1–9. https://doi.org/10.1098/rspb.2017.2770

Khayat, N., & Hochstein, S. (2019). Relating categorization to set summary statistics perception. *Attention, Perception, & Psychophysics*, *81*(8), 2850–2872. https://doi.org/10.3758/s13414-019-01792-7

Kleiner, M., Brainard, D. H., Pelli, D. G., Ingling, A., Murray, R., & Broussard, C. (2007). *What's new in Psychtoolbox-3*. Cognitive & Computational Psychophysics, Max Planck Institute for Biological Cybernetics.

Knoblich, G. (Ed.). (2006). *Human body perception from the inside out*. Oxford University Press.

Kohler, P. J., Caplovitz, G. P., & Tse, P. U. (2014). The global slowdown effect: Why does perceptual grouping reduce perceived speed? *Attention, Perception & Psychophysics*, *76*(3), 780–792. https://doi.org/10.3758/s13414-013-0607-x

Kriegeskorte, N. (2015). Deep neural networks: A new framework for modelling biological vision and brain information processing. *Annual Review of Vision Science*, *1*, 417–446. https://doi.org/10.1101/029876

Laloyaux, C., Destrebecqz, A., & Cleeremans, A. (2006). Implicit change identification: A replication of Fernandez-Duque and Thornton (2003). *Journal of Experimental Psychology: Human Perception and Performance*, *32*(6), 1366–1379. https://doi.org/10.1037/0096-1523.32.6.1366

Landy, M. S. (2014). Texture analysis and perception. In J. S. Werner & L. M. Chalupa (Eds.), *The New Visual Neurosciences* (pp. 639–652). MIT Press.

Levi, D. M. (2008). Crowding—An essential bottleneck for object recognition: A mini-review. *Vision Research*, *48*(5), 635–654. https://doi.org/10.1016/j.visres.2007.12.009

Manassi, M., & Whitney, D. (2018). Multi-level Crowding and the Paradox of Object Recognition in Clutter. *Current Biology*, *28*(3), R127–R133. https://doi.org/10.1016/j.cub.2017.12.051

Marchant, A. P., Simons, D. J., & de Fockert, J. W. (2013). Ensemble representations: Effects of set size and item heterogeneity on average size perception. *Acta Psychologica*, *142*(2), 245–250. https://doi.org/10.1016/j.actpsy.2012.11.002

Massey, D. S. (2002). A Brief History of Human Society: The Origin and Role of Emotion in Social Life. *American Sociological Review*, *67*(1), 1–29. https://doi.org/10.2307/3088931

Mather, G., & Parsons, T. (2018). Adaptation reveals sensory and decision components in the visual estimation of locomotion speed. *Scientific Reports*, *8*(13059), 1–8. https://doi.org/10.1038/s41598-018-30230-1

Mather, G., Radford, K., & West, S. (1992). Low-level visual processing of biological motion. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *249*(1325), 149–155. https://doi.org/10.1098/rspb.1992.0097

Mather, G., Sharman, R. J., & Parsons, T. (2017). Visual adaptation alters the apparent speed of real-world actions. *Scientific Reports*, *7*(1), 1–10. https://doi.org/10.1038/s41598-017-06841-5

Maule, J., & Franklin, A. (2016). Accurate rapid averaging of multihue ensembles is due to a limited capacity subsampling mechanism. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *33*(3), 22–29. https://doi.org/10.1364/JOSAA.33.000A22

McKee, S. P., & Welch, L. (1985). Sequential recruitment in the discrimination of velocity. *Journal of the Optical Society of America A*, *2*(2), 243–251. https://doi.org/10.1364/JOSAA.2.000243

Michalak, J., Troje, N. F., Fischer, J., Vollmar, P., Heidenreich, T., & Schulte, D. (2009). Embodiment of sadness and depression—Gait patterns associated with dysphoric mood. *Psychosomatic Medicine*, *71*(5), 580–587. https://doi.org/10.1097/PSY.0b013e3181a2515c

Moussaïd, M., Perozo, N., Garnier, S., Helbing, D., & Theraulaz, G. (2010). The walking behaviour of pedestrian social groups and its impact on crowd dynamics. *PloS One*, *5*(4), 1–7. https://doi.org/10.1371/journal.pone.0010047

Myczek, K., & Simons, D. J. (2008). Better than average: Alternatives to statistical summary representations for rapid judgments of average size. *Perception & Psychophysics*, *70*(5), 772–788. https://doi.org/10.3758/pp.70.5.772

Neri, P., Luu, J. Y., & Levi, D. M. (2006). Meaningful interactions can enhance visual discrimination of human agents. *Nature Neuroscience*, *9*(9), 1186–1192. https://doi.org/10.1038/nn1759

Neri, P., Morrone, M. C., & Burr, D. C. (1998). Seeing Biological Motion. *Nature*, *395*, 894–896.

Oriet, C., & Hozempa, K. (2016). Incidental statistical summary representation over time. *Journal of Vision*, *16*(3), 1–14. https://doi.org/10.1167/16.3.3

O'Toole, A. J., Roark, D. A., & Abdi, H. (2002). Recognizing moving faces: A psychological and neural synthesis. *Trends in Cognitive Sciences*, *6*(6), 261–266. https://doi.org/10.1016/S1364-6613(02)01908-3

Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., & Morgan, M. (2001). Compulsory averaging of crowded orientation signals in human vision. *Nature Neuroscience*, *4*(7), 739–744. https://doi.org/10.1038/89532

Pavlova, M. A. (2012). Biological Motion Processing as a Hallmark of Social Cognition. *Cerebral Cortex*, *22*(5), 981–995. https://doi.org/10.1093/cercor/bhr156

Pavlova, M. A., & Sokolov, A. (2000). Orientation specificity in biological motion perception. *Perception & Psychophysics*, *62*(5), 889–899. https://doi.org/10.3758/BF03212075

Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: Transforming numbers into movies. *Spatial Vision*, *10*(4), 437–442.

Pelli, D. G., & Tillman, K. A. (2008). The uncrowded window of object recognition. *Nature Neuroscience*, *11*(10), 1129–1135. https://doi.org/10.1038/nn.2187

Peng, S., Kuang, B., & Hu, P. (2019). Memory of Ensemble Representation Was Independent of Attention. *Frontiers in Psychology*, *10*, 1–8. https://doi.org/10.3389/fpsyg.2019.00228

Petzschner, F. H., Glasauer, S., & Stephan, K. E. (2015). A Bayesian perspective on magnitude estimation. *Trends in Cognitive Sciences*, *19*(5), 285–293. https://doi.org/10.1016/j.tics.2015.03.002

Pollick, F. E., Paterson, H. M., Bruderlin, A., & Sanford, A. J. (2001). Perceiving affect from arm movement. *Cognition*, *82*(2), 51–61. https://doi.org/10.1016/S0010-0277(01)00147-0

Rensink, R. A. (2000). The Dynamic Representation of Scenes. *Visual Cognition*, *7*(1–3), 17–42. https://doi.org/10.1080/135062800394667

Solomon, J. A., Morgan, M., & Chubb, C. (2011). Efficiencies for the statistics of size discrimination. *Journal of Vision*, *11*(12), 13. https://doi.org/10.1167/11.12.13

Sweeny, T. D., Haroz, S., & Whitney, D. (2012). Reference repulsion in the categorical perception of biological motion. *Vision Research*, *64*, 26–34. https://doi.org/10.1016/j.visres.2012.05.008

Sweeny, T. D., Haroz, S., & Whitney, D. (2013). Perceiving group behavior: Sensitive ensemble coding mechanisms for biological motion of human crowds. *Journal of*

Experimental Psychology. Human Perception and Performance, *39*(2), 329–337. https://doi.org/10.1037/a0028712

Sweeny, T. D., & Whitney, D. (2014). Perceiving crowd attention: Ensemble perception of a crowd's gaze. *Psychological Science*, *25*(10), 1903–1913. https://doi.org/10.1177/0956797614544510

Thompson, J., & Parasuraman, R. (2012). Attention, biological motion, and action recognition. *NeuroImage*, *59*(1), 4–13. https://doi.org/10.1016/j.neuroimage.2011.05.044

Thornton, I. M. (2012). Top-Down Versus Bottom-Up Processing of Biological Motion. In K. Johnson & M. Shiffrar (Eds.), *People Watching: Social, Perceptual, and Neurophysiological Studies of Body Perception* (pp. 25–43). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195393705.003.0003

Thornton, I. M., Rensink, R. A., & Shiffrar, M. (2002). Active versus passive processing of biological motion. *Perception*, *31*(7), 837–853. https://doi.org/10.1068/p3072

Thornton, I. M., & Vuong, Q. C. (2004). Incidental processing of biological motion. *Current Biology*, *14*(12), 1084–1089. https://doi.org/10.1016/j.cub.2004.06.025

Thornton, I. M., Vuong, Q. C., & Mather, G. (2019). Influence of Crowd Behaviour on Estimates of Biological Motion Speed. *Perception*, *48*(1_suppl), 36–36. https://doi.org/10.1177/0301006618824879

Thornton, I. M., Wootton, Z., & Pedmanson, P. (2014). Matching biological motion at extreme distances. *Journal of Vision*, *14*(3), 1–18. https://doi.org/10.1167/14.3.13

Treisman, A. (2006). How the deployment of attention determines what we see. *Visual Cognition*, *14*(4–8), 411–443. https://doi.org/10.1080/13506280500195250

Troje, N. F. (2002). Decomposing biological motion: A framework for analysis and synthesis of human gait patterns. *Journal of Vision*, *2*, 371–387. https://doi.org/10.1167/2.5.2

Troje, N. F., & Westhoff, C. (2006). The inversion effect in biological motion perception: Evidence for a 'life detector'? *Current Biology: CB*, *16*(8), 821–824. https://doi.org/10.1016/j.cub.2006.03.022

Tudor-Locke, C., Han, H., Aguiar, E. J., Barreira, T. V., Schuna, J. M., Kang, M., & Rowe, D. A. (2018). How fast is fast enough? Walking cadence (steps/min) as a practical estimate of intensity in adults: a narrative review. *British Journal of Sports Medicine*, *52*(12), 776–788. https://doi.org/10.1136/bjsports-2017-097628

Ueda, H., Yamamoto, K., & Watanabe, K. (2018). Contribution of global and local biological motion information to speed perception and discrimination. *Journal of Vision*, *18*(3), 1–11. https://doi.org/10.1167/18.3.2

Vanrie, J., & Verfaillie, K. (2004). Perception of biological motion: A stimulus set of human point-light actions. *Behavior Research Methods, Instruments, & Computers: A Journal of the Psychonomic Society, Inc*, *36*(4), 625–629. https://doi.org/10.3758/bf03206542

Ward, E. J., Bear, A., & Scholl, B. J. (2016). Can you perceive ensembles without perceiving individuals?: The role of statistical perception in determining whether awareness overflows access. *Cognition*, *152*, 78–86. https://doi.org/10.1016/j.cognition.2016.01.010

Watamaniuk, S. N., & Duchon, A. (1992). The human visual system averages speed information. *Vision Research*, *32*(5), 931–941. https://doi.org/10.1016/0042-6989(92)90036-i

Watamaniuk, S. N., & Heinen, S. J. (1999). Human smooth pursuit direction discrimination. *Vision Research*, *39*(1), 59–70. https://doi.org/10.1016/s0042-6989(98)00128-x

Watamaniuk, S. N., & McKee, S. P. (1998). Simultaneous encoding of direction at a local and global scale. *Perception & Psychophysics*, *60*(2), 191–200. https://doi.org/10.3758/bf03206028

Webster, J., Kay, P., & Webster, M. A. (2014). Perceiving the average hue of color arrays. *Journal of the Optical Society of America. A, Optics, Image Science, and Vision*, *31*(4), 283–292. https://doi.org/10.1364/JOSAA.31.00A283

Whitney, D., Haberman, J., & Sweeny, T. D. (2014). From Textures to Crowds: Multiple Levels of Summary Statistical Perception. In J. S. Wener & L. M. Chalupa (Eds.), *The New Visual Neurosciences* (pp. 695–710). MIT Press.

Whitney, D., & Levi, D. M. (2011). Visual crowding: A fundamental limit on conscious perception and object recognition. *Trends in Cognitive Sciences*, *15*(4), 160–168. https://doi.org/10.1016/j.tics.2011.02.005

Whitney, D., & Yamanashi Leib, A. (2018). Ensemble Perception. *Annual Review of Psychology*, *69*, 105–129. https://doi.org/10.1146/annurev-psych-010416-044232

Williams, D. W., & Sekuler, R. (1984). Coherent global motion percepts from stochastic local motions. *Vision Research*, *24*(1), 55–62. https://doi.org/10.1016/0042-6989(84)90144-5

Yamanashi Leib, A., Kosovicheva, A., & Whitney, D. (2016). Fast ensemble representations for abstract visual impressions. *Nature Communications*, *7*(13186), 1–10. https://doi.org/10.1038/ncomms13186

Yovel, G., & O'Toole, A. J. (2016). Recognizing People in Motion. *Trends in Cognitive Sciences*, *20*(5), 383–395. https://doi.org/10.1016/j.tics.2016.02.005